Timing Attention: from Reaction Time to Models of Visual Attention

Dissertation
"kumulativ"
zur Erlangung des Grades einer
Doktorin der Naturwissenschaften
(Dr.rer.nat.)
des Fachbereichs Psychologie der
Philipps-Universität Marburg

Vorgelegt von
Neda Meibodi
aus Borujerd, Iran

Marburg, 2022

Vom Fachbereich Psychologie
der Philipps-Universität Marburg (Hochschulkennziffer 1180)
als Dissertation angenommen am 2.12.2022

Erstgutachter: Prof. Dr. Dominik Endres
Zweitgutachterin: Prof. Dr. Anna Schubö

Tag der Disputation: 24.03.2023

Abstract

Models of visual attention have been widely proposed over the last two decades. Researchers in different disciplines, such as psychology and engineering, are interested in these models in order to understand human perceptual mechanisms and/or build algorithms which mimic the attentional processes for some applications (e.g. robotics).

In this dissertation I modeled the effect of learning experiences on attentional guidance. The presented model is an algorithmic-level model which links display inputs to the participants' reaction times. This dissertation consists of three studies.

In **the first study** the role of selection history –as the effect of learning from the practice phase of the experiment on the main phase– is investigated. I also tested dimension-level (e.g. color and shape) and feature-level (e.g. blue and red) selection histories. The results showed the version of the model which includes selection history (on feature-level), beside stimulus-driven (bottom-up) and goal-driven (top-down) control mechanisms, is best suited for a quantitative description of the participants' reaction times.

In **the second study**, I investigated the importance of intertrial priming –the effect of a previous trial on the current one– as well as the importance of each feature map (color, shape or orientation) in the model predictions. It was shown that by including the effect of intertrial priming a better description of the behavioral database can be achieved. Additionally, excluding any of the feature maps deteriorates the model predictions.

In **the third study** I proposed a model to decompose reaction times –into

decision and sensorimotor components– as a prerequisite of RT modeling. This study will help us introduce more accurate attention models. Furthermore, it can support cognitive studies to better investigate the effect of certain factors (e.g. age and mental disorders) on motor system vs. decision making.

The proposed attention model (in the first and the second study) is one of the first models that includes the selection history effect on guiding attention. This model can capture the between-group differences where each group of participants had a different learning experience. The model considers total reaction times of each participant. But attention can influence reaction times by affecting different cognitive processes. The third study introduces a method which helps us look at each process (and its relevant reaction time component) independently.

Acknowledgement

My PhD time was a life-changing experience. I would like to thank everyone who helped and supported me throughout these years. First and foremost, I would like to deeply thank my supervisor Prof. Dominik Endres for all that I have learned from him, for all interesting meetings and discussion we had, for his belief in me, and most importantly for being always supportive.

I am also grateful to have Prof. Anna Schubö as my second supervisor. Her valuable comments guided me to see my research from a different perspective. She was always ready to help, regardless how tight the deadlines were.

A very special thank to the other members of the thesis committee, Prof. Alexander Schütz and Prof. Hanna Christiansen for accepting my invitation to witness the last part of my PhD journey.

Thanks should also go to my great colleagues in the 'Theoretical Cognitive Science' lab for their kind support during my PhD. Beside our great work relationship I especially value our gatherings and chats in our lunch breaks and on our cake days. A special thank you to my colleague, Dr. Benjamin Knopp, for reviewing my thesis and his helpful comments.

I would like to deeply thank my coauthor and good friend, Hossein Abbasi, not only for his comments on papers and this thesis but also for his fantastic support in all first steps of my PhD life. Another special thanks to my close friend Zahra Khosrowtaj for always being by my side in both sweet and bitter moments of my life in Marburg.

More importantly, thank to my parents and my siblings in Iran for their generous love and to my sweet nieces, Adrina and Avina, for keeping my heart warm. Last but not the least, thanks to Andreas Wessner for all his support in the last year of my PhD which made this path easier and more pleasant for me and also many thanks to him for proofreading the thesis.

CONTENTS

CHAPTER 1

INTRODUCTION

Attention is so present in any single moment of our daily life that it is hard to imagine human beings without it. Visual attention, specifically, is one of the most fundamentals of cognition for all creatures with vision as the dominant sense (Moore & Zirnsak, 2017). Driving to a bakery, selective visual attention helps us to navigate, drive safely and finally find our favorite bread roll among many similar options. This ability to filter whatever is not needed in order to pay attention to whatever is needed not only gives taste to our life (selecting a yummy croissant) but also rescues us while driving. But how can all these complicated processes work while facing a vast variety of sensory inputs? This topic still has many open questions despite several decades of experimental studies, theoretical ideation and computational modeling. In most scenarios, we don't know what the brain exactly does with these sensory inputs but since these processes occur over time, one chance has been measuring the time and analyzing reaction times. For over a century, time

metric has been part of the psychophysical research (Luce, 1986). Naturally the metric of reaction times has been used to study the topic of attention. (Nobre & Coull, 2010).

The main motivation of this dissertation was modeling the connection between sensory inputs and the output –in this case reaction time (RT) probability distributions– for selective visual attention when different attentional mechanisms are in the game. We successfully modeled this connection by implementing some well-known attention theories and machine learning methods. In this chapter, I review the concepts which help the reader to follow the manuscripts and give a general understanding about the dissertation topic.

1.1 Selective visual attention

Selective visual attention is an ability that allocates brain processing resources to focus on important information of a scene (Zhang & Lin, 2013). Some important views on how this process works are described in the following sections.

1.1.1 Attentional mechanisms

Selective visual attention is defined as an ability to abandon whatever is irrelevant to catch the relevant items. Basically, this might happen due to voluntary intentions (**top-down**) or uniqueness of a stimulus (**bottom-up**). Bottom-up (stimulus-driven) and top-down (goal-driven) mechanisms and their race for attention capturing have been widely studied. Later **selection**

history was introduced as the third category of the attentional mechanisms. Selection history (Awh, Belopolsky, & Theeuwes, 2012; Theeuwes & Failing, 2020) or prior history (Wolfe & Horowitz, 2017; Wolfe, 2019) is a type of attentional control which comes from previous experience and can be organized by time scale, from within a trial to lifetime learning (Wolfe, 2021).

In a study Feldmann-Wüstefeld, Uengoer, and Schubö (2015) investigated selection history effect on attention deployment in four experiments. The result of this study proved the long term effect of selection history on attention deployment and also its power to cancel the predominant top-down control. In this study participants had to do a 'learning' task and a 'search' task (See Fig. 1.1). These tasks were mixed in the same block in the first and second experiment. They were performed in two consequent blocks within one day in experiment three, and over two days in experiment four. In the learning task (Fig. 1.1 a), participants categorized either color of the color singleton (green vs. blue) or shape of the shape singleton (triangle vs. pentagon). For half of the participants the color and for the other half the shape was response-relevant. In the search task a new shape singleton (diamond) was the target. In half of the search trial a color singleton (red) was present as a distractor (Fig. 1.1 b). The result of the experiments showed the participants' learning experience influenced attention. So, the participants who categorized color were more distracted by the red color singleton than the other group of participants. The experiment was carried out in two phases, a 'practice' and a 'main' phase. In the practice phase, participants had to learn either the color or the shape was the response-relevant dimension by responding to learning trials. In the main phase, search and learning trials

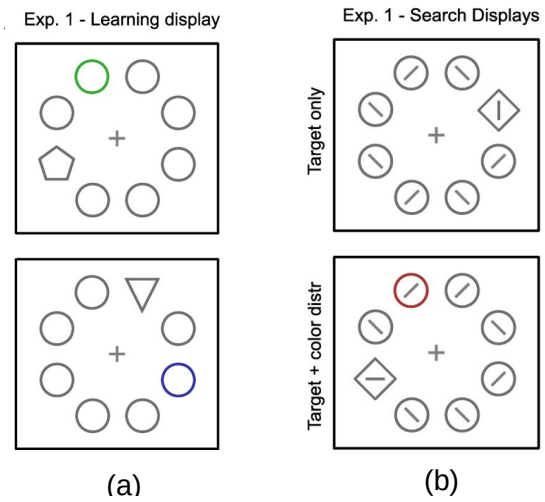

Figure 1.1: Experimental displays from (Feldmann-Wüstefeld et al., 2015). See the text for more information.

were randomly mixed and participants had to respond to both tasks in one experimental block.

Another well-known form of the history effect is **intertrial priming** (Theeuwes & van der Burg, 2011): an attentional bias toward a target feature which was selected in the very recent past (e.g. last trial). Theeuwes and van der Burg (2011) indicated that intertrial priming affects the saliency of a feature within the priority map and thus changes the selection priority. Moreover, Liesefeld, Liesefeld, Pollmann, and Müller (2019) claimed that intertrial priming effect is largely dimension-specific rather than feature-specific. In other words, repeating a specific feature (e.g. blue) can not improve participants' performance beyond repeating the whole feature dimension (e.g. color dimension) (Liesefeld et al., 2019).

It is commonly believed that the attentional mechanisms (top-down, bottom-up and selection history) feed into an integrated **priority map** (Theeuwes, Bogaerts, & van Moorselaar, 2022). Fig 1.2 depicts this pro-

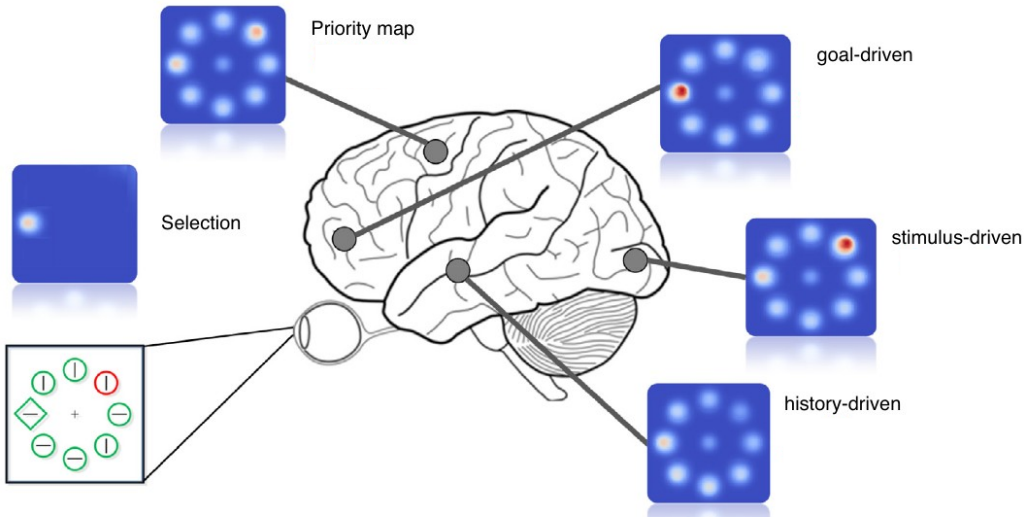

Figure 1.2: Schematic of the priority map, adapted from (Theeuwes, 2019). See also (Klink et al., 2014) for the basic version. The figure depicts how three maps (the goal-driven (top-up) map in frontal cortex, the stimulus-driven saliency (bottom-up) map in visual cortex and the history-driven map in the midbrain) are combined to highlight a stimuli and lead attention toward it.

cess and involved brain regions. A priority map is a winner-take-all neural mechanism that guides the allocation of covert and overt attention (Awh et al., 2012; Klink, Jentgens, & Lorteije, 2014). Within this map the mentioned mechanisms are combined in a weighted manner which determines the selection priorities (Theeuwes, 2019). The priority map is continuously available and updated during a task search (Wolfe, 2021).

1.1.2 Attention theories

Feature-based attention:

Feature integrated theory (FIT) is a two-stage, parallel-serial dichotomy on visual selective attention (Treisman & Gelade, 1980). As one of the most successful theories of attention, Treisman's FIT has been applied in several dis-

ciplines (e.g. psychology and neuroscience) for the last 40 years (Kristjánsson & Egeth, 2020; Bichot, Heard, DeGennaro, & Desimone, 2015). In the **preattentive** stage of FIT, stimulus features such as color, size and orientation are encoded into feature maps in parallel (see Fig.1.3). In addition to feature maps, FIT has a master map of locations that indicates which features are present at each spot of the display. At the second stage of attentive processing a set of features which are located at a selected position on the master map are integrated. This integration leads to recognition and localization of the object (Sanders, 1998). The need to attend to objects in order to recognize them, raises some problems: Firstly, in absence of attention the features might be floating and then conjunct illusorily (Kristjánsson & Egeth, 2020). Secondly, it is unclear how many items visual attention can process in a given short time and how attention is deployed to a target in a reasonable amount of time (Kristjánsson & Egeth, 2020; Wolfe & Horowitz, 2017). The answer might be that some guidance mechanisms e.g. bottom-up and top-down combine and guide attention to a subset of available objects (Wolfe & Horowitz, 2017).

Guided Search 2.0:

The human search engine has both serial and parallel search tasks. While FIT proposed a dichotomy between parallel and serial search tasks, Wolfe (1994) suggested that these tasks lie on a continuum with guided search tasks in the middle. As an example (see Fig. 1.4), if your target is a red 'T' which is among some other red and black letters, there is no need to search through all the letters and you just need to search through red ones. In

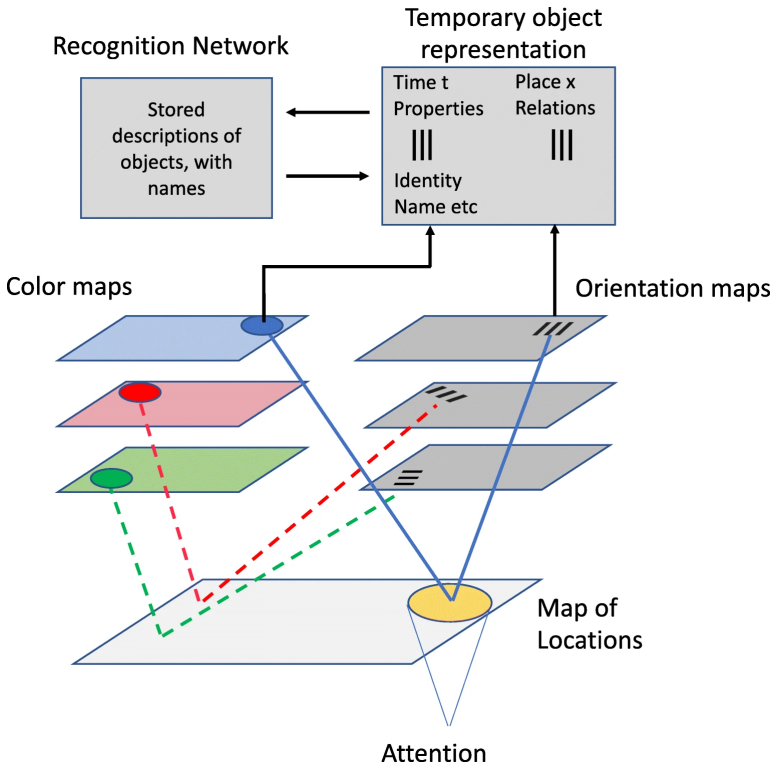

Figure 1.3: Treisman's feature integration theory. Figure from (Kristjánsson & Egeth, 2020). See the text for description.

other words, you use the knowledge about the basic features of the target to guide your attention toward items which have the chance of being the target. This was called Guided Search (GS) (Wolfe, 2014, 1994). Guided Search 2.0 (Fig. 1.5), the most known version of GS, argued that bottom-up maps can be modulated by top-down commands. These mechanisms combine –in a weighted manner– in a priority map and attention is guided to the highest peak of the priority map. In this process attention does not return to the distractors by implementing 'inhibition of return'.

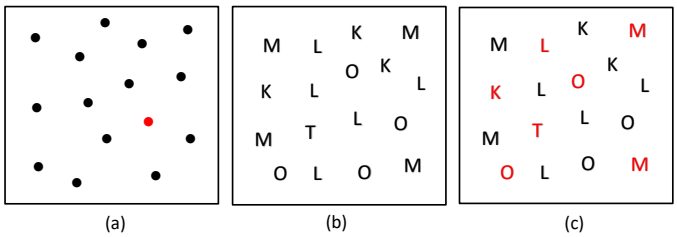

Figure 1.4: Search tasks (see the discussion in (Wolfe, 2015)). Parallel search (a): the red item pops out. Serial search (b): if T is the target, a letter by letter search is required to identify it. Guided search (c): if T is the target, the search is accelerated if it is known that the target has the color red (a similar example is depicted and discussed in (Wolfe, 2014)).

Guided search 6.0:

The current version of Guided Search (GS 6.0) is presented as a major update of GS 2.0. In GS 6.0, Wolfe (2021) claimed that there are two pathways at work when you guide your attention to your target. One of them is the non-selective pathway (which brings information like gist and scene properties) and the other is selective pathway which has a selective bottleneck. Access to the bottleneck i.e. attentional selection is guided by a priority map (Wolfe, 2021). In GS 6.0 there are five types of guidance which combine preattentively in the priority map: bottom-up, top-down, reward, history and scene guidance.

1.1.3 Attention models

Bottom-up models of attention:

Bottom-up saliency is the most modeled aspect of visual guidance. A large number of traditional saliency models (more than 70) are listed in (Koehler, Guo, Zhang, & Eckstein, 2014) considering their task types (object selection,

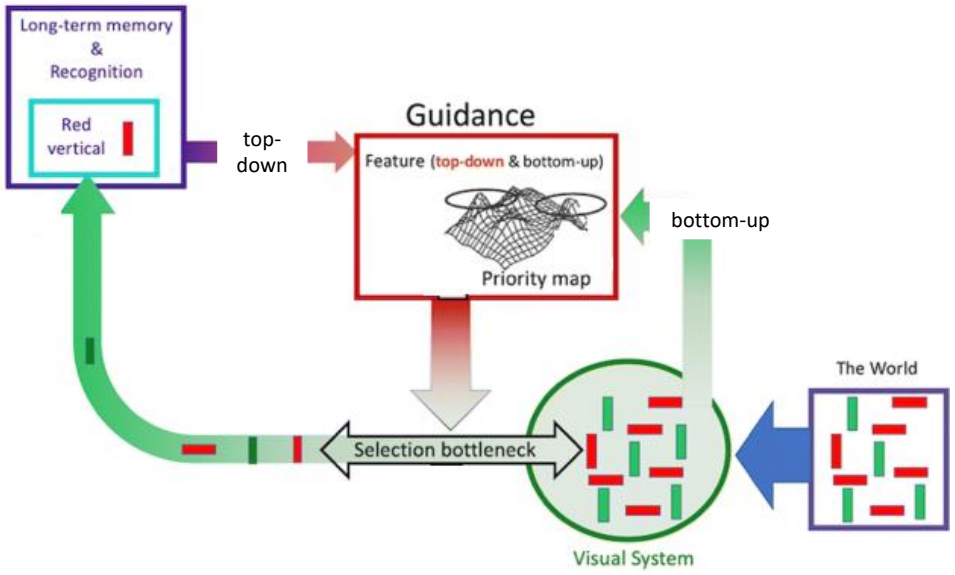

Figure 1.5: A schematic representation of Guided Search 2.0 adapted from (Wolfe, 2021).

free viewing and task-based viewing). In another review, Itti and Borji (2015) classified many saliency models, based on their biological or mathematical inspirations, into seven categories such as cognitive models (Itti, Koch, & Niebur, 1998), information-theoretic models (Bruce & Tsotsos, 2009) and graphical models (Zhang, Tong, Marks, Shan, & Cottrell, 2008).

Itti's model (Itti et al., 1998; Itti & Koch, 2000) is one of the most innovative models which implemented feature integration theory to predict saliency. This model has three feature maps including color, intensity and orientation. Within each map just locations, which stand out in their neighborhood, persist. These locations are fed into a master saliency map which codes the most noticeable locations over the whole visual scene. Finally the most salient location is found in a winner-take-all manner while the **inhibition of return** mechanism guarantees attention going from each attended

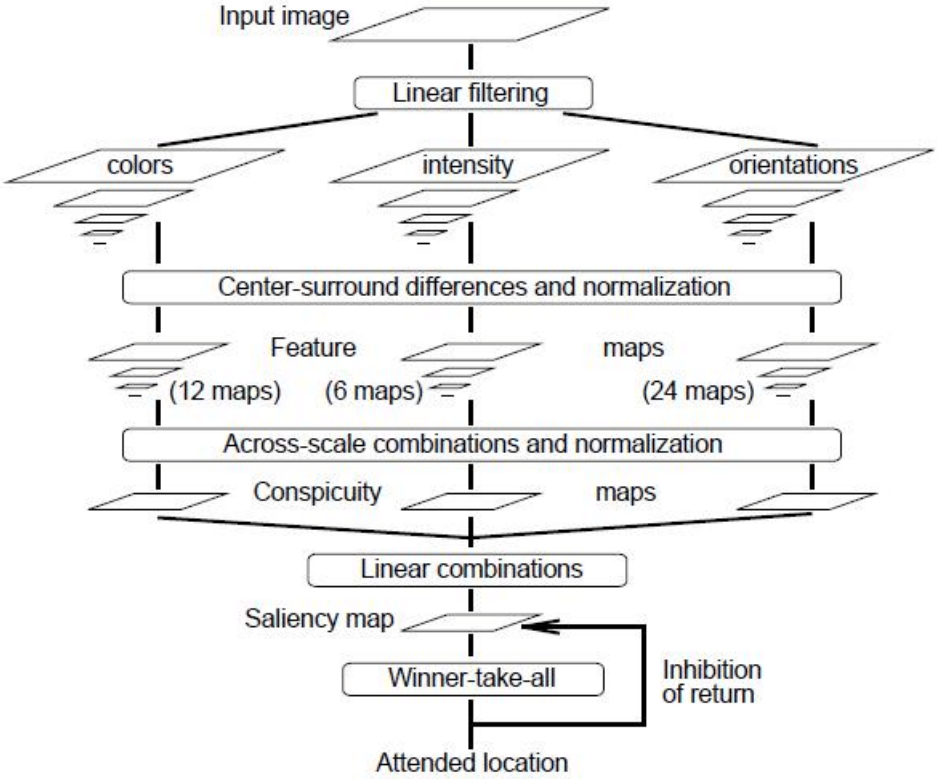

Figure 1.6: Saliency model architecture proposed by Itti and Koch (2000).

saliency to the next. Fig. 1.6 shows a schematic of this model. The model has been successfully implemented (de Brecht & Saiki, 2006; Veale, Hafed, & Yoshida, 2017), expanded (Tanner & Itti, 2019; Ramirez-Moreno, Schwartz, & Ramirez-Villegas, 2013) and reviewed (Koehler et al., 2014) over the last two decades.

Another state-of-the-art saliency model which is famously known as 'Attention based on Information Maximization (AIM)' was proposed by Bruce and Tsotsos (2009). AIM consists of two major parts (Zhang & Lin, 2013). First, the model is trained on a set of image patches using ICA (Independent Component Analysis). The patches are randomly sampled from 3600 natu-

ral images (Bruce, 2008). Second, the saliency at each point of an image is measured by Shannon's self-information of that location with respect to its surrounding context (Kim & Milanfar, 2013).

Both Itti and AIM models are compared in (Koehler et al., 2014) to discover in which type of tasks human behavior is best predicted by these saliency models. The results show the models cannot predict human eye movements in free viewing tasks although these models are successful in predicting explicit saliency. It seems that even when there is no instruction (such as free viewing tasks), observers make their own tasks allowing other attentional controls to come into play (Wolfe & Horowitz, 2017).

Saliency models have made major progress in performance in last decade thanks to utilizing deep learning algorithms and also the easier access to large image databases (Borji, 2019). Many researchers believe that deep neural networks might help us learn more about human visual search, especially where there is an ambiguity e.g. the role of shape in attention guidance (Wolfe & Horowitz, 2017). Furthermore, most of the traditional models like Itti's model cannot extract higher level features or objects while deep models have proved their ability in that (Borji, 2019). However, there is still a gap between human performance and deep models. For instance the models can not understand the high-level semantics such as action, text and unusuality in rich scenes (Borji, 2019). Additionally, pre-trained deep neural networks might not have an acceptable performance on laboratory search task images (Li, 2022). Fig. 1.7 shows two examples when a trained network (SALICON) neglects low-level image features although the network is successful in tracking high-level features, such as faces. More discussion around this topic can

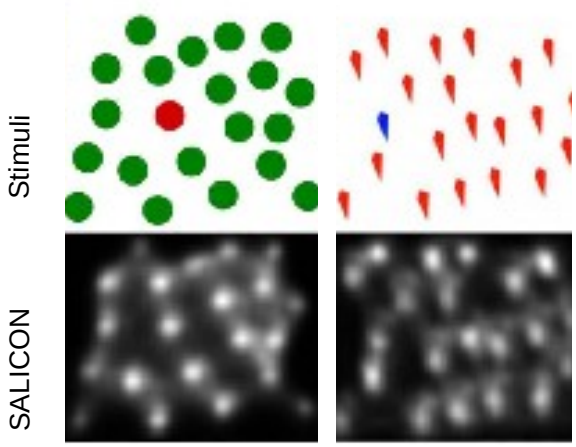

Figure 1.7: Deep neural network failure cases. These two examples show that a deep neural network (in this case SALICON) does not perform well in tracking local contrast. The bottom row shows predicted saliency maps by SALICON. The right side and the left side examples are adapted from (Huang et al., 2015) and (Rahman & Bruce, 2015), respectively.

be found in (Borji, 2019).

Integrated bottom-up and top-down models of attention:

Top-down attention is not modeled as extensively as bottom-up attention due to its complex nature (Tanner & Itti, 2019) or limitations such as being influenced by other mechanisms (e.g. intertrial priming (Theeuwes & van der Burg, 2011)). Nevertheless, it is essential to propose models including more mechanisms since attention deployment is the result of the competition between several factors. Consequently, in the last two decades, a vast amount of presented models focused on including both bottom-up and top-down mechanisms (Itti & Borji, 2015). Itti's bottom-up model was implemented in a top-down model (Tanner & Itti, 2019) to represent the effect of goal-relevance information on attention and eye-movement.

Bayesian models of attention have shown how bottom-up noisy sensory input and top-down priors can be hierarchically integrated by cortical neurons to reduce uncertainty in guiding attention to a target (Rao, 2005). Under the assumption that probabilities represent the neurons' firing rates, Chikkerur, Serre, Tan, and Poggio (2010) mapped a Bayesian network's nodes to the nodes of a biological model to show the role of attention –involving the cortical ventral and dorsal pathways– in answering to 'what is where' questions. Borji and Itti (2013) claimed that these Bayesian models can explain more complex attention mechanisms over time which gives them good prediction power. But their drawbacks lie in model complexity, especially when it comes to training (Borji & Itti, 2013).

1.2 Reaction time

Since using mechanical machines such as Galton's pendulum (Dodonova & Dodonov, 2013) to this computer era, reaction time (RT) has been captivating researchers in psychophysical studies. RTs can help us understand cognitive processes although they need to be used with caution since they might be influenced by many factors such as fatigue (van den Berg & Neely, 2006), age (Woods, Wyma, Yund, Herron, & Reed, 2015), gender (Dykiert, Der, Starr, & Deary, 2012), physical activities (Jain, Bansal, Kumar, & Singh, 2015) or computer hardware and software (Dodonova & Dodonov, 2013). Many 'simple reaction time' (SRT) studies have measured the effect of these factors on observers' response times. SRT is the time between the appearance of a known stimulus and receiving the response (e.g. a pressed

button).

1.2.1 Reaction time distributions

Psychophysical studies mostly only focus on the mean of RT, however the advantages of analyzing the full distribution has been pointed out by many researchers. Palmer, Horowitz, Torralba, and Wolfe (2011) mentioned that mean analyzing –which comes from the assumption of RT normality– might lead to imprecise conclusions. For example, excluding outliers based on their distance from the mean in terms of standard deviations might mistakenly cause the informative part of RTs being removed (RT distributions are not normal and are skewed to the right side).

Using RT distributions can also be important for modeling purposes –as it is in the models presented in this thesis. Many studies showed that distributions which have an exponential component (e.g. ex-Gaussian) fit better on RTs (Palmer et al., 2011). In our studies we tested several distributions and found the best fit (ex-Gaussian followed by an inverse Gaussian) for our data. See these two distributions in Fig. 1.8 and compare with Gaussian distribution in the same figure. More examples can be found in the second manuscript. We needed the distribution which best describes our data because our model predicts the parameters of the data distribution as the posterior. Below, I review ex-Gaussian and inverse Gaussian distributions. The second one is also used in the third study to model the decision component of RT.

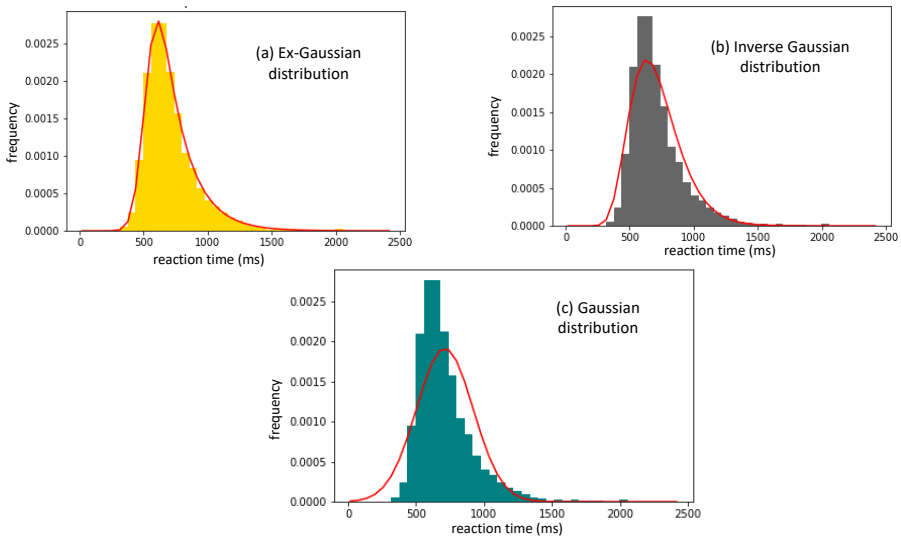

Figure 1.8: Reaction time distributions: ex-Gaussian (a), inverse Gaussian (b) and Gaussian (c). These distributions are fitted on the data (from (Feldmann-Wüstefeld et al., 2015)) by maximizing the log-likelihoods.

Ex-Gaussian distribution:

The ex-Gaussian distribution is a convolution of Gaussian and exponential distributions, see e.g. (Luce, 1986). It has three parameters: μ, σ and τ that are the mean and standard deviation of the Gaussian component and the mean of the exponential component, respectively. The mean and the variance of this distribution are $\mu+\tau$ and $\sigma^2+\tau^2$. Equation (1.1) shows the probability density function (Moret-Tatay, Gamermann, Navarro-Pardo, & Castellá, 2018) where erfc is the complementary error function:

$$f(t) = \frac{1}{2\tau} e^{\frac{1}{2\tau}(2\mu+\frac{\sigma^2}{\tau}-2t)} \text{erfc}\left(\frac{\mu+\frac{\sigma^2}{\tau}-t}{\sqrt{2}\sigma}\right) \tag{1.1}$$

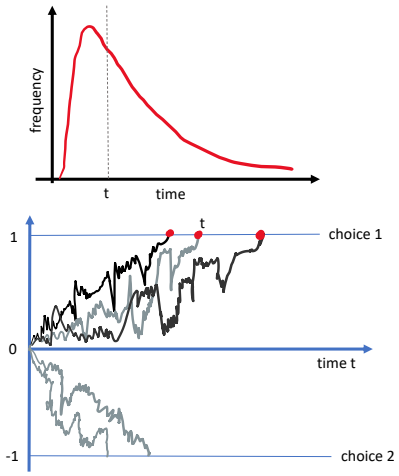

Figure 1.9: Graphical example of an evidence accumulation model for a two-alternative forced choice task. It is proven that the first passage time has an inverse Gaussian distribution (see the text).

Inverse Gaussian distribution:

If RT is the time needed for an evidence accumulation to reach to a fixed boundary, it is distributed as an inverse Gaussian. See Fig. 1.9.

The first passage time distribution can be obtained by supposing $W(t)$ is a Wiener process in one dimension with positive drift ν and variance σ^2, and that $W(0) = 0$. Then T, the time required for $W(t)$ to reach the value a for the first time, is a random variable with a density function (Folks & Chhikara, 1978) which is an inverse Gaussian distribution:

$$f(t;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi t^3}} \exp\left(-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right),\ \lambda = \frac{a^2}{\sigma^2},\ \mu = \frac{a}{\nu} \tag{1.2}$$

1.2.2 Reaction time components

It is believed that the total observed reaction time (RT) is the sum of different time components, e.g. sensory delays, decision making and motor execution

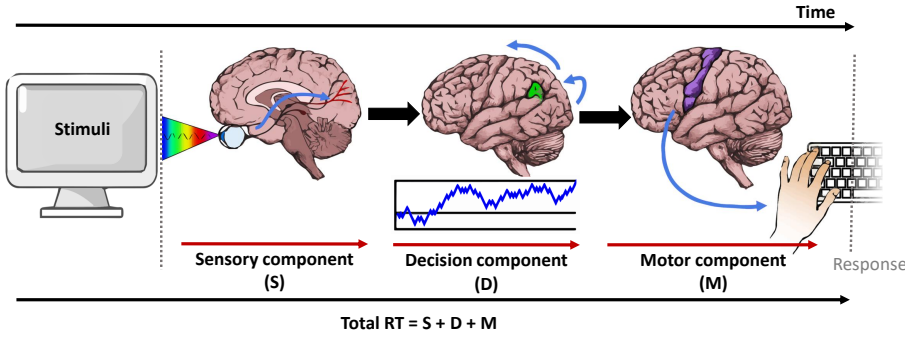

Figure 1.10: Components of reaction time (RT). Total RT is the sum of these components. The components can still be subdivided into smaller parts. The components may be overlapped at some points which is not considered in this figure. This figure is adapted from (*Mental chronometry*, n.d.).

(see Fig. 1.10). Any of these components can be subdivided further (Luce, 1986). In mental chronometry there have been many efforts to measure each process separately (Posner, 2005). The assumption that these processes occur in serial is not convincing because some components may be overlapping (Luce, 1986). For instance it is conceivable that decision process and motor components are intertwined (Evans & Wagenmakers, 2020). However, in the absence of enough evidence most researchers assumed the total RT is the sum of serial stages. In Ratcliff and Childers (2015)'s drift-diffusion model the non-decision component –the time that is spent on processes other than the decision making– happens before and after the decision part. Generally, how these components combine to yield final RTs' distributions, has been a matter of question for several decades. In some older research (Christie & Luce, 1956; McGill, 1963; Hohle, 1965) it is reported that RT is a convolution of a Gaussian and an exponentially distributed component (which results an ex-Gaussian distribution), where one represents the decision time and

another represents the motor component. The above mentioned authors had opposite believes about the source of this exponential component: Christie and Luce (1956) mentioned that decision time is exponentially distributed, but in contrast McGill (1963) related that to motor response. Schwarz (2001) proposed that total RT is a convolution of inverse Gaussian (Wald) and an exponentially distributed component (which results an ex-Wald distribution). His motivation to use inverse Gaussian was that it models the first passage time distribution of a random walk (Folks & Chhikara, 1978). Nevertheless, it has been doubted if an exponential distribution can be applied to either the decision or the motor component (Palmer et al., 2011).

Dissecting total observed RT into sub-stages enables to determine at which stage(s) a new factor (e.g., nicotine, sleep deprivation, or Parkinson's disease) had its influence (Posner, 2005). Another advantage can be modelling purposes e.g. attention models proposed in the first and the second study of this dissertation. This was my motivation to perform the third study when we decomposed RTs into decision and sensorimotor components. For the details see next chapter and also the manuscripts in appendixes.

CHAPTER 2

SUMMARY OF MANUSCRIPTS

In this chapter I summarize the manuscripts. The complete versions of the manuscripts are attached to this dissertation in the appendix.

2.1 First Manuscript

Neda Meibodi, Hossein Abbasi, Anna Schubö, & Dominik Endres (2021). **A model of selection history in visual attention**. Proceedings of the Annual Meeting of the Cognitive Science Society, 43. Retrieved from https://escholarship.org/uc/item/3m33h9h7

In this study I have modeled selective visual attention in presence of three attentional mechanisms, namely stimulus-driven (bottom-up), goal-driven (top-down) and selection history. The model links the sensory inputs to participants' reaction times. Fig. 2.1 illustrates how the model works. First, display inputs are encoded into the feature maps at the preattentive stage. Subsequently, the self-information of each feature map ($-log(p(f))$

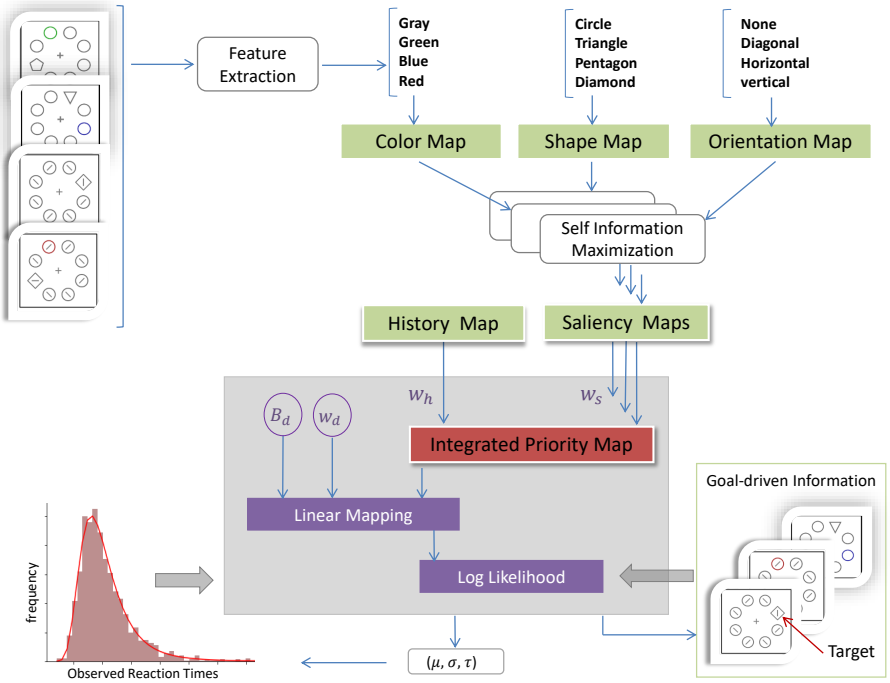

Figure 2.1: Schematic of the model published in the first manuscript (Meibodi et al., 2021). The blue arrows show the data flow direction. The grey arrows show the feedback from goal-relevant information and RTs to the machine learning part of the model (the grey box). Goal-relevant information helps the model to guide attention to the target location. w_s is the saliency weight with three elements for color, shape and orientation. w_h is the history map weight. w_d is the distribution parameters weight and has three elements for μ, σ and τ (ex-Gaussian distribution parameters). B_d is distribution parameters' prior containing B_μ, B_σ and B_τ.

is computed and yields a saliency map. This method was applied by Bruce and Tsotsos (2009) in a saliency model. Self-information maximization measures the local contrast and finds the most informative parts of the input i.e. surprise. Beside saliency maps, the effect of learning experience is also modelled by creating a 'history map' (more information can be found in the next paragraphs). These maps are integrated into a priority map in a weighted manner. The weights consist of w_h (history weight) and w_s (saliency weight)

as depicted in Fig. 2.1. Attention is guided to a location which has the highest activation in the priority map. The activation of each map equals the value of the map multiplied by the weight of the map.

In a Bayesian framework, we assumed that there is a linear mapping from the integrated priority map to the RT distributions. By means of this linear mapping the model machine-learns to predict the weights (w_h and w_s) and also the RT distribution parameters (B_d) for each participant. We visualized and quantified the closeness of the model-predicted distribution to the best-fit distribution for each participants' RTs. These results prove that the model predicts the RT distributions parameters (B_d) very accurately. Importantly, the maps weights (w_h and w_s) analysis confirmed the history effect on attention guidance.

The RT database of the first experiment proposed by Feldmann-Wüstefeld et al. (2015) is used in this study. See the experiment displays in Fig.1.1 and the text related to that in the previous chapter. The behavioural data of the main phase is modelled in this study. The model addresses the selection history as the influence of the practice-phase learning on the main phase of the experiment. We tested three versions of the model: the first one with feature-level selection history, the second one with dimensional-level selection history and the last one without history map. The result of the model comparison showed that the first model with feature-level selection history is the best suited one. In this version of the model, history map contains the response-relevant features in the practice phase (blue and green for the color group, triangle and pentagon for the shape group). In the second version of the model, it is assumed that the participants had learned to predict

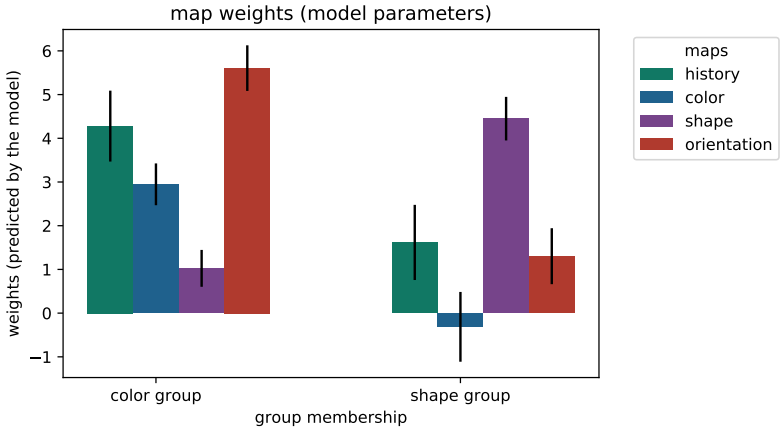

Figure 2.2: Saliency and history maps weights for both the color and shape groups (Meibodi et al., 2021). These weights are shown with w_s and w_h in Fig. 2.1. Note that w_s has three elements for color, shape and orientation. The error bars show the standard errors.

responses on the dimensional level (color or shape) and not on the level of single features (e.g. green or blue). Consequently, red (for color group participants) and diamond (for shape group participants) are also included in the history map.

The model captures between-group differences by giving different weights to the maps of the color and shape group participants. See these weights in Fig. 2.2 of the best version of the model (feature-level selection history). This figure shows that the 'history map' weight is higher in the color group than in the shape group. This shows that the color group participants had to rely more on their learning experience to solve the tasks. These participants learned about colors (blue and green) during the practice phase of the experiment which is seen (in the model) in the large weight of the history map. Although these colors are in the 'color map' too, there is another color (red) in this map which needs to be suppressed because it is a task-irrelevant

feature. This results in a smaller weight of the color map in comparison to the weight of the history map. In shape group, the color map is close to zero since this group had to ignore the color singletons in both tasks. Instead, the shape group focused on the 'shape map' which is response-relevant in both tasks. The weight of the 'orientation map' is larger in the color group than in the shape group, indicating that the color group relied on orientation saliency to respond to the search task.

The uniqueness of this model is its ability to process three attentional mechanisms: stimulus-driven, goal-driven and especially selection history which has not been modelled extensively in previous studies. However, the intertrial priming effect is not captured in the presented model. Exploring its influence on the created model was the main motivation of my second study.

2.2 Second Manuscript

Neda Meibodi, Hossein Abbasi, Anna Schubö, & Dominik Endres (submitted, 2022). **Distracted by Previous Experience: Integrating Selection History, Current Task Demands and Saliency in an Algorithmic Model**.

The motivation of this study was to expand the visual attention model presented in the first manuscript. Fundamentals of the model, created in the first study, were not altered. Please see Fig. 2.3 and compare with Fig. 2.1. In this study, the intertrial priming effect (as another form of selection history) (Theeuwes & van der Burg, 2011) is added to the existing model. We claim that intertrial priming emphasizes the response-relevant feature dimension of the last trial on the current one (Theeuwes & Failing, 2020). For instance, if a blue color singleton is response-relevant in trial (n-1), this can bias attention toward a different color singleton (e.g. green) in trial (n).

We tested different versions of the model to better understand the influence of each saliency map on the model. Each version excludes either one of the color, shape or orientation maps. In order to compare all versions of the model, we computed a Laplace-approximation (Bishop, 2006) to the Bayesian model evidence across all participants. The results are shown in Fig. 2.4. The model comparison results indicated that all maps including history, color, shape and orientation, as well as the intertrial priming effect are necessary to reach the best approximation of the RT database.

For the best version of the model (which is called M1 in Fig. 2.4), the map

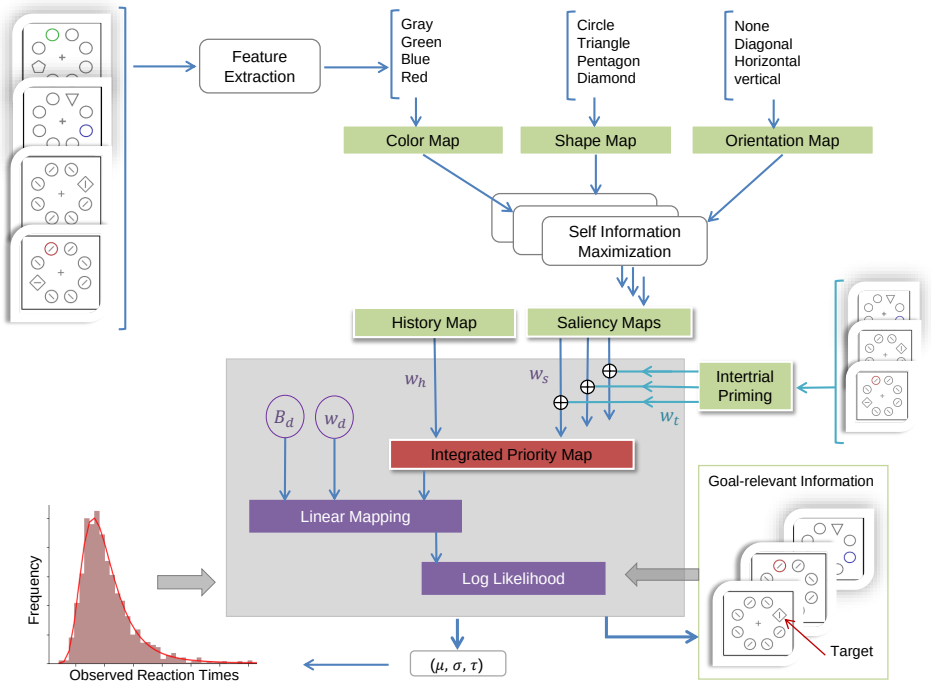

Figure 2.3: Schematic of the model published in the second manuscript (Meibodi et al., 2022). Intertrial priming is added to the model (right side). w_t is the intertrial priming weight and has three elements for color, shape and orientation.

weights are plotted to show the influence of intertrial priming on the weights. See these weights in Fig. 2.5 and compare with Fig. 2.2. Two points are very noticeable on this figure. Firstly, the high reliance of shape group participants on the shape priming which is explicit since the shape was response-relevant in both tasks for this group of participants. Secondly, in the color group, the intertrial priming effect increases the 'orientation map' weight and decreases the 'shape map' weight. This might indicate that switching from the search task (reporting orientation embedded in a shape singleton target) to the learning task (reporting colors and not shape singletons) is best managed by lowering the shape feature's weight and boosting the orientation weight.

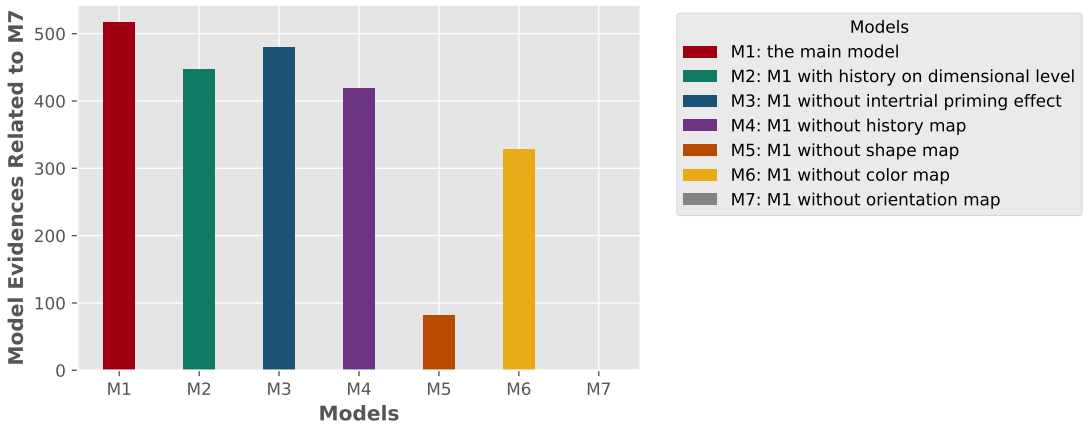

Figure 2.4: Model comparison (Meibodi et al., 2022). The model evidences are plotted relative to M7 (the least probable model). The main model (M1) which scores the best, comprises of the saliency maps (color, shape and orientation), the history map on feature-level (as it is explained in the first manuscript summary) and also the intertrial priming on dimension-level.

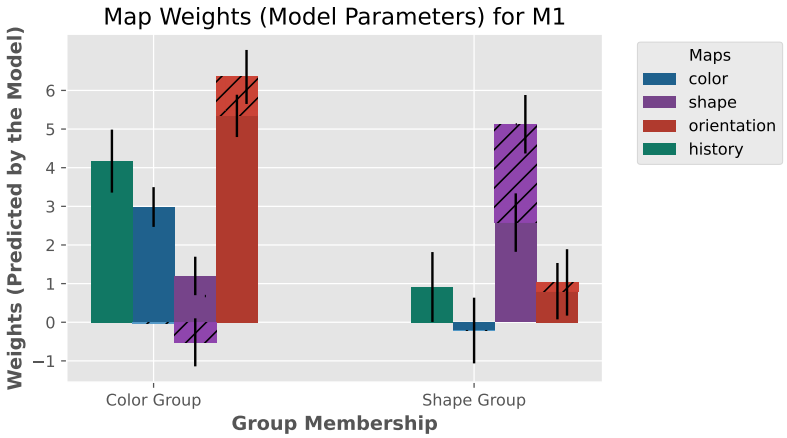

Figure 2.5: Map weights of both the color and the shape groups (Meibodi et al., 2022). The hatched parts are the intertrial priming weights. Note that the final weight of each saliency map is a sum of the saliency map weight (w_s) and the priming weight (w_t). Priming modulations on color maps are close to zero and can be hardly seen. The error bars represent the standard errors.

2.3 Third Manuscript

Neda Meibodi, Anna Schubö, & Dominik Endres (2022). **Sensorimotor processes are not a source of much noise: sensorimotor and decision components of reaction times**. Proceedings of the Annual Meeting of the Cognitive Science Society, 44. Retrieved from https://escholarship.org/uc/item/1nj6m2n7

The presented model disentangles RT distributions into two main components, 'decision time' and 'sensorimotor time', by using machine learning methods.

We assumed that the decision component can be viewed as the first passage time in a Wiener diffusion process. Thus, the distribution of the decision component is an inverse Gaussian (Schwarz, 2001). The distribution of the sensorimotor component has no strong theoretical evidence. We therefore tested Gaussian, gamma or Laplace distributions, mainly because of their popularity in the literature (Christie & Luce, 1956; Hohle, 1965; Ratcliff & Tuerlinckx, 2002) and their special shape (i.e. gamma distribution has a positive support or Laplace has heavier tails). We evaluated which model assumptions maximize approximate Bayesian model evidence (free energy (Friston, Kilner, & Harrison, 2006) or evidence lower bound (Bishop, 2006)). The results show that the version of the model which has Gaussian distribution as a sensorimotor component scores the best. For this version of the model, Fig. 2.6 visualizes how the subtraction of the sensorimotor component from the total RTs can improve the data-fitting of the inverse Gaussian distribution.

We also modeled the outliers. An outlier is either a very fast or a very

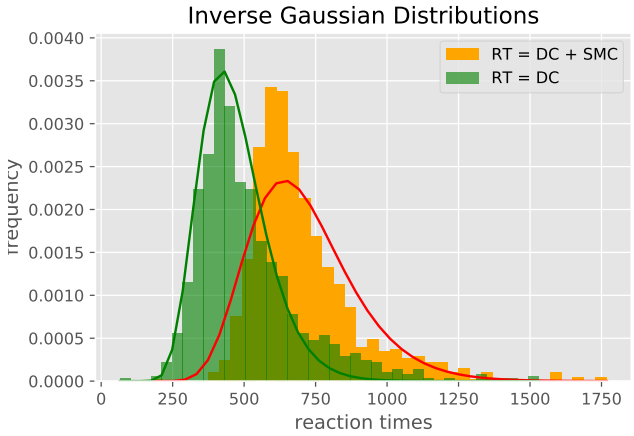

Figure 2.6: Inverse Gaussian distributions of a participant's RTs. The orange histogram contains the total RTs, red curve is fitted by maximizing log-likelihood. The green histogram shows the expected decision components (DC) after subtracting the sensorimotor component (SM). The parameters of the green distribution are predicted by the model. More examples can be found in (Meibodi et al., 2021).

slow response. The model assumption is that the outliers are uniformly distributed. All versions of the model label a very similar proportion of the trials as outliers, independent of the choice of sensorimotor distribution. More importantly, this outlier labeling is driven by the model assumption and does not need any additional criteria. Fig. 2.7 illustrates which part of a RT distribution (for one of the participants) is marked by the model as outliers.

The predicted mean of sensorimotor distribution is in the range of 199.58± 0.37 (ms) which is comparable to the reported mean in several 'simple reaction time' experiments (Amini Vishteh, Mirzajani, Jafarzadehpour, & Darvishpour, 2019; Jain et al., 2015). Note that the model predicted distributions are very narrow which was also assumed in (Ratcliff & Childers, 2015).

Identically to the first and second manuscript the RT database of (Feldmann-

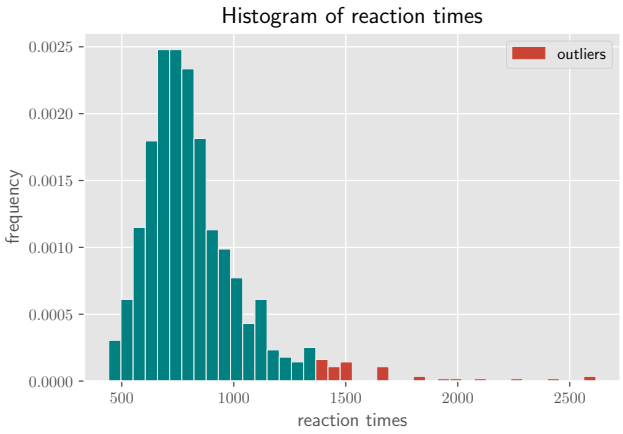

Figure 2.7: Outliers predicted by the model. These outliers are marked in red. More examples can be found in (Meibodi et al., 2021).

Wüstefeld et al., 2015) was used in this manuscript to optimize our model.

CHAPTER 3

DISCUSSION

Models of visual attention attempt to mimic, explain or predict parts or all of human visual attentive behavior (Tsotsos & Rothenstein, 2011). These models can be roughly classified into descriptive (Wolfe, 2021; Treisman & Gelade, 1980), neurobiological (Parr & Friston, 2017), computational (Bruce & Tsotsos, 2009) and algorithmic (Koch & Ullman, 1985) models. In general, visual attention modeling may have various purposes, such as image classification (Mnih, Heess, Graves, & Kavukcuoglu, 2014), computer vision and robotics (Hiruma, Mori, Ito, & Ogata, 2022; Frintrop, 2006) or studying a specific experimental observation (Meibodi et al., 2021). Consequently, comparing models does not seem straightforward, fair, or useful so it might be better to compare some parts of the models which have relevant functionalities (Tsotsos & Rothenstein, 2011).

In the first and the second studies of this dissertation, I presented an 'algorithmic model' of visual attention. According to Tsotsos and Rothen-

stein (2011)'s definition, algorithmic models "provide mathematics and algorithms that govern their performance and as a result present a process by which attention might be computed and deployed". In other words, they are a combination of 'descriptive' models and 'data fitting'. Data fitting models capture parameter variations in experimental data and show how well the models fit to the experimental data (Tsotsos & Rothenstein, 2011).

The presented model shows how several attention theories, such as feature integration theory, work in a united framework to link sensory inputs to behavioral outputs. The model considers three different attentional factors –bottom-up (stimulus-driven), top-down (goal-driven) and selection history– which makes this model more comprehensive than previously reported attention models. The model successfully predicts the RT distribution parameters for each participant while weighting the saliency and history maps. The value of these weights differ based on the participants' learning history and targets.

Our model is similar to GS 2.0 (Wolfe, 1994) in the way that it describes attention as a result of bottom-up and top-down activation in the priority map. Unlike our model, GS is a descriptive presentation of attention and has not been implemented on an experimental dataset. In the current version of this model, GS 6.0 (Wolfe, 2021), more attentional factors (reward, scene information and prior history) are taken into consideration. Clearly, a model with more mechanisms will give a better chance to move toward future naturalistic models since attention in the real world is influenced by many factors and not just saliency. We have already built the model composing three factors. Including more mechanisms in this model requires future experimental and modeling plans.

From the model's limitations to future plans:

1. Although we have learned a lot from lab-designed experiments, this is the time to move toward semi-naturalistic and naturalistic experiments and models. To accomplish this, the model must be able to handle the information which is expected in real –non lab– environments. Firstly, the model already extracts simple features (color, shape and orientation) at the preattentive stage which may not be sufficient to cope with more complex inputs. Secondly, although the proposed model integrates three attentional factors, a naturalistic model might need to incorporate more factors, such as the scene structure.

2. One of the model's assumptions is that the participants' selection history comes from a learning-practice phase. Our model captures participants' behavior after this reinforcement learning phase is completed. Therefore, our model does not include an explicit reinforcement learning component. This could be an intriguing aspect to further investigate and incorporate into the model.

3. The model is one of the first selection history models, and needs to be extended in the future by adding more attentional dynamics aspects. We need to progress toward a dynamic model in which both the selection history and the dynamic priority map are updated over time.

4. In this study (see the first and the second manuscript), I modeled total RT. The relativity of different RT components to different cognitive processes was not considered. To resolve this shortcoming, we proposed a model (see the third manuscript) which decomposes the total

observed RT into decision and sensorimotor components. My model will be useful, not only in updating our attention models, but also whenever RT components need to be extracted in cognitive RT modeling. This model should be tested on another dataset to see the effect of the stimuli features (e.g. color or intensity) on the reported mean and on the shape of the sensorimotor component. As a possible future plan I am interested in applying the model on datasets from different psychophysical studies to see the effect of certain parameters such as age or disorders e.g. ADHD (Pedersen, Frank, & Biele, 2017) and Parkinson (Herz, Bogacz, & Brown, 2016; Low, Miller, & Vierck, 2002) on different RT components.

References

Amini Vishteh, R., Mirzajani, A., Jafarzadehpour, E., & Darvishpour, S. (2019). Evaluation of simple visual reaction time of different colored light stimuli in visually normal students. *Clinical Optometry*, *11*, 167–171. doi:10.2147/OPTO.S236328

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. doi:10.1016/j.tics.2012.06.010

Bichot, N. P., Heard, M. T., DeGennaro, E. M., & Desimone, R. (2015). A Source for Feature-Based Attention in the Prefrontal Cortex. *Neuron*, *88*(4), 832–844. Retrieved from `http://dx.doi.org/10.1016/j.neuron.2015.10.001` doi:10.1016/j.neuron.2015.10.001

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Borji, A. (2019). *Saliency prediction in the deep learning era: Successes, limitations, and future challenges.* Retrieved from `https://arxiv.org/abs/1810.03716`

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 185–207. doi:10.1109/TPAMI.2012.89

Bruce, N. D. B. (2008). *Saliency , attention , and visual search: An information theoretic approach*. Unpublished doctoral dissertation.

Bruce, N. D. B., & Tsotsos, J. K. (2009). Saliency , attention , and visual search: An information theoretic approach. *Journal of Vision*, *9*(3), 1–24. doi:10.1167/9.3.5.

Chikkerur, S., Serre, T., Tan, C., & Poggio, T. (2010). What and where: A Bayesian inference theory of attention. *Vision Research*, *50*(22), 2233–2247. doi:10.1016/j.visres.2010.05.013

Christie, L., & Luce, R. D. (1956). Decision structure and time relations in simple choice behavior. *The bulletin of mathematical biophysics*, *18*, 89–112.

de Brecht, M., & Saiki, J. (2006). A neural network implementation of a saliency map model. *Neural Networks*, *19*(10), 1467–1474.

doi:10.1016/j.neunet.2005.12.004

Dodonova, Y. A., & Dodonov, Y. S. (2013). Is there any evidence of historical slowing of reaction time? No, unless we compare apples and oranges. *Intelligence*, *41*(5), 674–687. doi:10.1016/j.intell.2013.09.001

Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). Sex differences in reaction time mean and intraindividual variability across the life span. *Developmental Psychology*, *48*(5), 1262–1276. doi:10.1037/a0027550

Evans, N. J., & Wagenmakers, E.-J. (2020). Evidence Accumulation Models: Current Limitations and Future Directions. *The Quantitative Methods for Psychology*, *16*(2), 73–90. doi:10.20982/tqmp.16.2.p073

Feldmann-Wüstefeld, T., Uengoer, M., & Schubö, A. (2015). You see what you have learned. Evidence for an interrelation of associative learning and visual selective attention. *Psychophysiology*, *52*(11), 1483–1497. doi:10.1111/psyp.12514

Folks, J. L., & Chhikara, R. S. (1978). The Inverse Gaussian Distribution and Its Statistical Application-A Review. *Journal of the Royal Statistical Society*, *40*(3), 263–289. doi:10.1111/j.2517-6161.1978.tb01039.x

Frintrop, S. (2006). *VOCUS: A visual attention system for object detection and goal-directed search.* Unpublished doctoral dissertation. doi:10.1007/11682110

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology Paris*, *100*(1-3), 70–87. doi:10.1016/j.jphysparis.2006.10.001

Herz, D. M., Bogacz, R., & Brown, P. (2016). Neuroscience: Impaired Decision-Making in Parkinson's Disease. *Current Biology*, *26*(14), R671–R673. doi:10.1016/j.cub.2016.05.075

Hiruma, H., Mori, H., Ito, H., & Ogata, T. (2022). Guided Visual Attention Model Based on Interactions Between Top-down and Bottom-up Information for Robot Pose Prediction. Retrieved from `http://arxiv.org/abs/2202.10036`

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*(4), 382–386. doi:10.1037/h0021740

Huang, X., Shen, C., Boix, X., & Zhao, Q. (2015). SALICON: Reducing the semantic gap in saliency prediction by adapting deep neural networks. *Proceedings of the IEEE International Conference on Computer Vision*, 262–270. doi:10.1109/ICCV.2015.38

Itti, L., & Borji, A. (2015). *Computational models: Bottom-up and top-down aspects.* Retrieved from http://arxiv.org/abs/1510.07748 doi:10.48550/arXiv.1510.07748

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. , *40*, 1489–1506. doi:10.1016/S0042-6989(99)00163-7

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliancy-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254 – 1259. doi:10.1109/34.730558

Jain, A., Bansal, R., Kumar, A., & Singh, K. (2015). A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *International Journal of Applied and Basic Medical Research*, *5*(2), 124. doi:10.4103/2229-516x.157168

Kim, C., & Milanfar, P. (2013). Visual saliency in noisy images. *Journal of Vision*, *13*(4), 1–14. doi:10.1167/13.4.5

Klink, P. C., Jentgens, P., & Lorteije, J. A. M. (2014). Priority maps explain the roles of value, attention, and salience in goal-oriented behavior. *Journal of Neuroscience*, *34*(42), 13867–13869. doi:10.1523/jneurosci.3249-14.2014

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–227. doi:10.1007/978-94-009-3833-5

Koehler, K. L., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict? *Journal of Vision*, *14*(3), 1–27. doi:10.1167/14.3.14

Kristjánsson, Á., & Egeth, H. (2020). How feature integration theory integrated cognitive psychology, neurophysiology, and psychophysics. *At-*

tention, Perception, and Psychophysics, *82*, 7–23. doi:10.3758/s13414-019-01803-7

Li, Q. (2022). Understanding Saliency Prediction with Deep Convolutional Neural Networks and Psychophysical Models. Retrieved from http://arxiv.org/abs/2204.06071

Liesefeld, H. R., Liesefeld, A. M., Pollmann, S., & Müller, H. J. (2019). Biasing allocations of attention via selective weighting of saliency signals: behavioral and neuroimaging evidence for the Dimension-Weighting Account. In *In t. hodgson (ed.), processes of visuospatial attention and working memory.* Springer. doi:10.1007/7854_2018_75

Low, K. A., Miller, J., & Vierck, E. (2002). Response slowing in Parkinson's disease: A psychophysiological analysis of premotor and motor processes. *Brain*, *125*(9), 1980–1994. doi:10.1093/brain/awf206

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* Oxford University Press.

McGill, W. J. (1963). Stochastic latency mechanisms. *In D. Luce (ed.), Handbook of Mathematical Psychology, John Wiley & Sons.*, 1–309.

Meibodi, N., Abbasi, H., Schubö, A., & Endres, D. (2021). A model of selection history in visual attention. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*. Retrieved from https://escholarship.org/uc/item/3m33h9h7

Meibodi, N., Abbasi, H., Schubö, A., & Endres, D. (2022). Distracted by Previous Experience: Integrating Selection History, Current Task Demands and Saliency in an Algorithmic Model. *submitted*.

Mental chronometry. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Mental_chronometry

Mnih, V., Heess, N., Graves, A., & Kavukcuoglu, K. (2014). Recurrent models of visual attention. *Advances in Neural Information Processing Systems*, *3*(January), 2204–2212.

Moore, T., & Zirnsak, M. (2017). Neural Mechanisms of Selective Visual Attention. *Annual Review of Psychology*, *68*, 47–72. doi:10.1146/annurev-psych-122414-033400

Moret-Tatay, C., Gamermann, D., Navarro-Pardo, E., & Castellá, P. F. d. C.

(2018). ExGUtils: A python package for statistical analysis with the ex-Gaussian probability density. *Frontiers in Psychology*, *9*(612), 1–11. doi:10.3389/fpsyg.2018.00612

Nobre, A. C., & Coull, J. T. (2010). *Attention and Time*. Oxford University Press. Retrieved from https://doi.org/10.1093/acprof:oso/9780199563456.001.0001

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 58–71. doi:10.1037/a0020747

Parr, T., & Friston, K. J. (2017). The active construction of the visual world. *Neuropsychologia*, *104*, 92–101. doi:10.1016/j.neuropsychologia.2017.08.003

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin and Review*, *24*(4), 1234–1251. doi:10.3758/s13423-016-1199-y

Posner, M. I. (2005). Timing the brain: Mental chronometry as a tool in neuroscience. *PLoS Biology*, *3*(2), 0204–0206. doi:10.1371/journal.pbio.0030051

Rahman, S., & Bruce, N. D. (2015). Saliency, scale and information: Towards a unifying theory. *Advances in Neural Information Processing Systems*, 2188–2196. Retrieved from https://proceedings.neurips.cc/paper/2015/file/a51fb975227d6640e4fe47854476d133-Paper.pdf

Ramirez-Moreno, D. F., Schwartz, O., & Ramirez-Villegas, J. F. (2013). A saliency-based bottom-up visual attention model for dynamic scenes analysis. *Biological Cybernetics*, *107*(2), 141–160. doi:10.1007/s00422-012-0542-2

Rao, R. P. N. (2005). Bayesian inference and attentional modulation in the visual cortex. *COGNITIVE NEUROSCIENCE AND NEUROPSYCHOLOGY*, *16*(16), 1–6. Retrieved from papers2://publication/uuid/27A7ECBD-5AB3-4E33-999D-7B7B216754D0

Ratcliff, R., & Childers, R. (2015). Individual Differences and Fitting Meth-

ods for the Two-Choice Diffusion Model of Decision Making. *Decision*, *2*(4), 237–279.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review*, *9*(3), 438–481. doi:10.3758/BF03196302

Sanders, A. F. (1998). Elements of human performance: Reaction processes and attention in human skill.

Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, and Computers*, *33*(4), 457–469. doi:10.3758/BF03195403

Tanner, J., & Itti, L. (2019). A top-down saliency model with goal relevance. *Journal of Vision*, *19*(1), 1–16. doi:10.1167/19.1.11

Theeuwes, J. (2019). Goal-driven, stimulus-driven, and history-driven selection. *Current Opinion in Psychology*, *29*, 97–101. doi:10.1016/j.copsyc.2018.12.024

Theeuwes, J., Bogaerts, L., & van Moorselaar, D. (2022). What to expect where and when: how statistical learning drives visual selection. *Trends in Cognitive Sciences*, *26*(10), 860–872. doi:10.1016/j.tics.2022.06.001

Theeuwes, J., & Failing, M. (2020). *Attentional Selection: Top-Down, Bottom-Up and History-Based Biases (Elements in Perception)*. Cambridge: Cambridge University Press. doi:10.1017/9781108891288

Theeuwes, J., & van der Burg, E. (2011). On the limits of top-down control of visual selection. *Attention, Perception, and Psychophysics*, *73*(7), 2092–2103. doi:10.3758/s13414-011-0176-9

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. doi:10.1016/0010-0285(80)90005-5

Tsotsos, J. K., & Rothenstein, A. (2011). *Computational models of visual attention.* doi:10.4249/scholarpedia.6201

van den Berg, J., & Neely, G. (2006). Performance on a simple reaction time task while sleep deprived. *Percept Mot Skills*, *102*(2), 589–599. doi:10.2466/pms.102.2.589-599. PMID: 16826680

Veale, R., Hafed, Z. M., & Yoshida, M. (2017). How is visual salience computed in the brain? Insights from behaviour, neurobiology and modeling. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *372*(1714), 1–14. doi:10.1098/rstb.2016.0113

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Wolfe, J. M. (2014). Approaches to Visual Search: Feature Integration Theory and Guided Search. *The Oxford Handbook of Attention.*, 1–40. doi:10.1093/OXFORDHB/9780199675111.013.002

Wolfe, J. M. (2015). *The Human Visual Search Engine.* Retrieved from https://serious-science.org/human-visual-search-engine-1942

Wolfe, J. M. (2019). Visual Attention: The Multiple Ways in which History Shapes Selection. *Current Biology*, *29*(5), R155–R156. Retrieved from https://doi.org/10.1016/j.cub.2019.01.032 doi:10.1016/j.cub.2019.01.032

Wolfe, J. M. (2021). *Guided Search 6.0: An updated model of visual search* (Vol. 28) (No. 4). Psychonomic Bulletin & Review. doi:10.3758/s13423-020-01859-9

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1–8. doi:10.1038/s41562-017-0058

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience*, *9*. doi:10.3389/fnhum.2015.00131

Zhang, L., & Lin, W. (2013). *Selective Visual Attention: Computational Models and Applications.* John Wiley & Sons, Ltd. doi:10.1002/9780470828144

Zhang, L., Tong, M. H., Marks, T. K., Shan, H., & Cottrell, G. W. (2008). SUN: A Bayesian framework for saliency using natural statistics. *Journal of Vision*, *8*(7), 1–20. doi:10.1167/8.7.32

APPENDIX A

AUTHOR CONTRIBUTIONS

1. Neda Meibodi (70%), Hossein Abbasi (10%), Anna Schubö (10%), & Dominik Endres (10%) (2021). **A model of selection history in visual attention**. Proceedings of the Annual Meeting of the Cognitive Science Society, 43. Retrieved from https://escholarship.org/uc/item/3m33h9h7

2. Neda Meibodi (70%), Hossein Abbasi (10%), Anna Schubö (10%), & Dominik Endres (10%) (submitted, 2022). **Distracted by Previous Experience: Integrating Selection History, Current Task Demands and Saliency in an Algorithmic Model**.

3. Neda Meibodi (80%) , Anna Schubö (10%), & Dominik Endres (10%) (2022). **Sensorimotor processes are not a source of much noise : sensorimotor and decision components of reaction times**. Proceedings of the Annual Meeting of the Cognitive Science Society, 44. Retrieved from https://escholarship.org/uc/item/1nj6m2n7

Prof. Dr. Dominik Endres .....

Neda Meiboi .....

APPENDIX B

FIRST MANUSCRIPT

A model of selection history in visual attention

Neda Meibodi (meibodi@uni-marburg.de)

Hossein Abbasi (hossein.abbasi@uni-marburg.de)

Anna Schubö (schuboe@uni-marburg.de)

Dominik Endres (dominik.endres@uni-marburg.de)

Department of Psychology, Philipps-University Marburg, Gutenbergstrasse 18,
35032 Marburg, Germany

Abstract

Attention can be biased by the previous learning and experience. We present an algorithmic-level model of this bias in visual attention that predicts quantitatively how bottom-up, top-down and selection history compete to control attention. In the model, the output of saliency maps as bottom-up guidance interacts with a history map that encodes learning effects and a top-down task control to prioritize visual features. We test the model on a reaction-time (RT) data set from the experiment presented in (Feldmann-Wüstefeld, Uengoer, & Schubö, 2015). The model accurately predicts parameters of reaction time distributions from an integrated priority map that is comprised of an optimal, weighted combination of separate maps. Analysis of the weights confirms learning history effects on attention guidance.

Keywords: Visual attention; Selection history; Integrated priority map; Self information maximization; Feature integrated theory; Ex-Gaussian distribution

Introduction

Selective visual attention is a brain function that filters irrelevant sensory inputs to facilitate focusing on relevant items. Bottom-up and top-down mechanisms have traditionally been proposed to control the process of attention guidance. Object saliency and environment features shape the attentional process in a bottom-up manner while the top-down process is mostly controlled by observer intentions and preferences.

In addition to top-down and bottom-up contributions also 'selection history' can play a significant role in guiding attention toward a specific target (Theeuwes, 2019). Selection history (as a third category of attentional deployment) comes into play when an object is emphasized just because of previous attendance in the same context (Awh, Belopolsky, & Theeuwes, 2012). To clarify the distinction between top-down guidance and selection history, Theeuwes argued that selection history is a fast, effortless, and automatic version of attention control while top-down selection is slow, effortful, and controlled (Theeuwes, 2018).

One special form of selection history has been investigated in (Feldmann-Wüstefeld et al., 2015; Kadel, Feldmann-Wüstefeld, & Schubö, 2017; Henare, Kadel, & Schubö, 2020). These studies combined an associative learning task with a visual search task. The result showed that observers attend more to a stimulus which was predictive in the preceding feature discrimination task. Considering to what extent selection history can be suppressed by top-down process, Kadel et al. (2017) tested three different top-down-influenced modes of task preparations such as pretrial task cuing. As their results showed, attentional biases induced by selection history persisted despite the task preparation.

An integrated priority map was proposed by Awh et al. as a theoretical framework to explain how selection history and other factors of attention guidance interact (Awh et al., 2012; Theeuwes, 2019). Priority maps have been successfully employed by many authors (Fecteau & Munoz, 2006; Zelinsky & Bisley, 2015; Klink, Jentgens, & Lorteije, 2014; Todd & Manaligod, 2017; Veale, Hafed, & Yoshida, 2017; Chelazzi et al., 2014) to explain the result of the processes which shape attention. In a review, Klink et al. (2014) summarized how goal-driven and stimulus-driven maps in cortex combine with a value-based map in midbrain. This combination results in a priority map for the frontal eye fields.

Stimulus-driven (bottom-up) models of attention were developed early on (Itti, Koch, & Niebur, 1998). These models tend to ignore the effects of selection history, task or training (Itti & Borji, 2015). Itti et al. (1998) implemented feature integration theory (three feature maps including color, intensity and orientation), winner-take-all, inhibition of return and a normalization method to model visual attention in a bottom-up manner. Veale et al. (2017) validated a neural implementation of Itti's model. In another bottom-up model, Bruce and Tsotsos (2006, 2009) –using self information maximization ($-\log(p(x))$), where x is a feature – proposed a computational model of saliency that is called 'Attention based on Information Maximization (AIM)', because attention is attracted by surprising, i.e. potentially informative, regions of an image. Furthermore, thanks to deep learning advances, there has been recent progress in deep visual saliency models (Borji, 2019).

Beside above mentioned models, Itti and Borji (2015) reviewed more than 50 computational bottom-up models. They also reviewed some computational top-down models. Such models (Navalpakkam & Itti, 2005; Hwang, Higgins, & Pomplun, 2009; Borji, Sihite, & Itti, 2014) are less well researched than saliency models, which might be due to the fact that they require information not available from the stimulus. There are also some models on how bottom-up and top-down work together in attentional guidance (Chikkerur, Serre, Tan, & Poggio, 2010; Kimura et al., 2008). Chikkerur et al. (2010) used a Bayesian framework to explain how a combination of bottom-up and top-down attentional guidance work together

in cortex.

Despite substantial progress in building models of attention, there are still many open questions. Selection history has hardly been modeled. One exception is Tseng et al.'s model of the influence of inter-trial priming – a type of selection history effect – on attention guidance (Tseng, Glaser, Caddigan, & Lleras, 2014). They implemented a Ratcliff-type diffusion model (Ratcliff, 1978) for a 2-forced-choice task and showed that the history can affect Ratcliff diffusion model parameters.

In this paper we introduce an algorithmic-level model (in the sense of Marr (1982)) to show how bottom-up, top-down and selection history compete against each other to guide visual attention toward a specific target. By selection history here we mean the effect of learning from previous experience on the current task (see (Feldmann-Wüstefeld et al., 2015; Kadel et al., 2017; Henare et al., 2020)). The model comprises priority maps to integrate goal-driven, saliency-based and history-related biases in a winner-take-all manner. Bottom-up guidance, feature maps and subsequently saliency maps are made based on 'feature integration theory' (Treisman & Gelade, 1980) and 'self information maximization' (AIM) (Bruce & Tsotsos, 2009). To reflect the effect of selection history and learning in the model, a history map contributes to the integrated priority map. Finally, task-relevant information controls the map integration weights that generate predictions for responses and response times. These integration weights are our model for the top-down influences. We test this model on a behavioral database from an experiment by Feldmann-Wüstefeld et al. (2015). The model can predict the reaction time distribution parameters for each participant and also across the experimental groups. To find the best distribution of reaction times, several probability density functions are compared maximizing log-likelihood and the best fitting one – an ex-Gaussian distribution (Matzke & Wagenmakers, 2009)– is used in the model.

Materials and methods

Experiment

The data used in this study comes from the first experiment of Feldmann-Wüstefeld et al. (2015). They investigated the impact of associative learning on covert selective visual attention. The experiment consisted of a 'practice' and a 'main' phase, in which two types of tasks (learning and search) were performed. A central fixation cross was presented on the screen, which was then surrounded by eight different elements on an imaginary circle (Figure 1). 28 participants were divided randomly into 2 different groups, namely 'color group' and 'shape group'. They were first naive about their group membership, but had to learn it on a trial and error basis in the practice phase.

In the 'practice phase', participants had to learn that either color or shape was the response relevant dimension in this learning task (see Figure 1A). Members of the color group had to report the color of the color singleton (blue or green), whereas members of the shape group had to respond to the shape of the shape singleton (triangle or pentagon). They had to use their left hands to press one of two buttons that were placed on the left side of the response pad. Auditory feedback indicated whether they pressed the incorrect key.

In the 'main phase' a second visual search task was added, and participants performed both tasks in random order. In the search task (Figure 1B), all participants had to report the orientation of a line presented inside a diamond shape target. In half of the trials, a response-irrelevant red circle was presented as distractor. Participants used their right hand to press one of two buttons on the right side of the pad to indicate the line orientation (horizontal versus vertical).

The results of this study showed that the history of selection acquired in the learning task affected the participants' performance in the search task. Stimuli that were predictive of the relevant dimension in the learning task biased attention in the visual search task. The authors suggested that the participants' history of either shape or color selection in the practice phase had resulted in a selection history bias.

We presented a model of this selection history bias in the current study based on the behavioral data from the main phase, which comprises at total of 28672 trials across all participants. More details about the experiment can be found in (Feldmann-Wüstefeld et al., 2015).

The Algorithmic Model

Based on the theoretical considerations outlined in the introduction and a preliminary data analysis, we assembled an algorithmic-level model to explain how top-down and bottom-up influences competitively interact with visual selection history to guide attention toward a specific stimulus. The results of this preliminary analysis, that was aimed at determining experimental factors influencing responses and reaction times, are not shown here for space constraints. Inspired by the integrated priority maps in (Awh et al., 2012), we used a 'history map' reflecting the influence of selection history on current attention deployment, see Figure 2. Additionally, there is an overall saliency map for bottom-up influences. How these maps combine into an integrated priority map is controlled by the task in a top-down fashion. Fig-

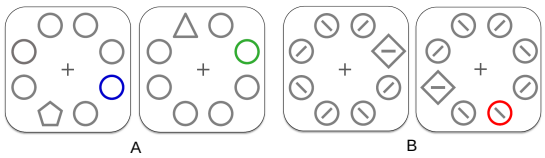

Figure 1: Learning task (A): Participants in the color group had to respond to the color (green vs. blue) and participants in the shape group had to respond to the shape (pentagon vs. triangle). Search task (B): The orientation (horizontal vs. vertical) of the line embedded in the diamond had to be reported. Distractor-absent trial (left). Distractor-present trial (right).

ure 2 also shows how the output of the integrated priority map feeds into a two-part neural network that predicts ex-Gaussian distribution parameters (Luce, 2008) of reaction times (left exit path in the figure) and response likelihoods (the right exit path).

The input stage of the model is based on feature-integration theory (Treisman & Gelade, 1980). The model extracts three types of features (color, shape and orientation) and feature maps –as shown in Figure 2– are computed. In the next processing step, saliency maps that model the effect of bottom-up control on visual attention (Koch & Ullman, 1985) are formed from the feature maps. Shannon's measure of Self-Information is applied, similar to Attention Based on Information Maximization (Bruce & Tsotsos, 2009), to compute saliency maps. Eq (1) and Eq (2) show the actual calculations behind map computation. Feature maps are $M \times N \times K$ vectors where M is the number of trials, N is the number of objects in each trial and K is the number of distinct values that each feature can take on, i.e. we are using 1-out-of-K encoding for the features, with the value 1 indicating which feature value is present. In the current experiment $M = 1024$ (for each participants), $N = 8$ and $K = 4$. Figure 3 illustrates the method of building feature maps for some example trials. For all trials, we take the feature maps f_i for $i \in \{color, shape, orientation\}$ and compute the self-information X_i:

$$\forall k : X_i[k] = -log\left(\sum_{n=1}^{N} f_i[n][k]/N\right) \tag{1}$$

which yields the saliency of all trials $s_i[n]$:

$$\forall n : s_i[n] = X_i\left[\arg\max_k(f_i[n][k])\right] \tag{2}$$

where, due to the 1-of-K feature encoding, we can use *argmax* to pick the self-information corresponding to the current feature value.

Saliency maps s_i are fed into the integrated priority map along with history information (h) to compete in a soft winner-take-all model (Theeuwes, 2019) for the predicted response target. Selection history, the third category of attentional guidance (Awh et al., 2012), carries the effect of learning (participants learned about color or shape in our experiment) into the priority map (p):

$$\forall m,n : p[m][n] = \underset{n}{\text{soft}\max}\left(\sum_i \left(w_{s_i} * s_i[m][n]\right) + w_h * h[m][n]\right) \tag{3}$$

The weights (w_h for history and w_{s_i} for $i \in \{color, shape, orientation\}$) are used to combine the history map and the saliency maps and reflect the effect of the task in a top-down manner. The softmax function is used to ensure that the winning location receives the most attention while keeping the map interpretable as a probability distribution. In our model, Eq 3 can be interpreted as the first layer of a (two-layer) neural network. The second layer is a (linear) mapping from the integrated priority map to reaction time distribution parameters:

$$\forall m : d = \sum_{n=1}^{N} (p[m][n] * w_d) + B_d \tag{4}$$

When w and B are weights and biases of ex-Gaussian distribution parameters' for $d[m] \in (\mu[m], \sigma[m], \tau[m])$.

We also compute a 1-out-of-K representation of the target information ($g[m][n]$ in Eq 5, see also Figure 3) which is used for machine-learning the weights with which the history map and the saliency maps are combined. The weights (w_h, w_{s_i} and w_d) for a task are determined by maximizing the log of the joint distribution of the reaction times (RT), the target g under the distribution predicted by the integrated priority map and the prior distributions over the model parameters δ:

$$\begin{aligned} L &= \sum_{m=1}^{M} \log\left(ExG\left(RT[m] \,|\, \mu[m], \sigma[m], \tau[m]\right)\right) \\ &+ \left(\sum_{m=1}^{M}\sum_{n=1}^{N} \log(p[m][n]) * g[m][n]\right) + \delta \end{aligned} \tag{5}$$

where ExG is ex-Gaussian distribution function. δ is computed as the sum of the logs of the following prior distributions:

$$\begin{aligned} w &\sim \mathcal{N}(0.0, 1.0) \\ B_\mu &\sim \mathcal{N}(600.0, 100.0) \\ B_{\sigma^2} &\sim \mathcal{N}(75.0, 4.0) \\ B_\tau &\sim \mathcal{N}(200.0, 20.0) \end{aligned} \tag{6}$$

Mean and standard deviation of these distributions are selected in a way that matches results from similar experiments (Feldmann-Wüstefeld et al., 2015; Kadel et al., 2017). To find the weights and biases that maximize the joint probability (Eq 5), we draw random initial values from these distributions and then optimize using Python 3.7.6, PyTorch 1.6.0 and Adam optimizer with learning rate 0.2. Code and training data for the models can be found here: `http://dx.doi.org/10.17192/fdr/64.2`

Results and Discussion

To investigate how selection history quantitatively influences attentional guidance, three versions of the model with different history maps are tested. In the first version, the history map contains the response-relevant features in the learning phase (blue and green for the color group, triangle and pentagon for the shape group). In the second version of the model, the history map includes all color singletons (for participants in the color group) and all shape singletons (for participants in the shape group). The assumption is that the participants have learned response-predictiveness on the dimensional level (color or shape), not on the level of single features

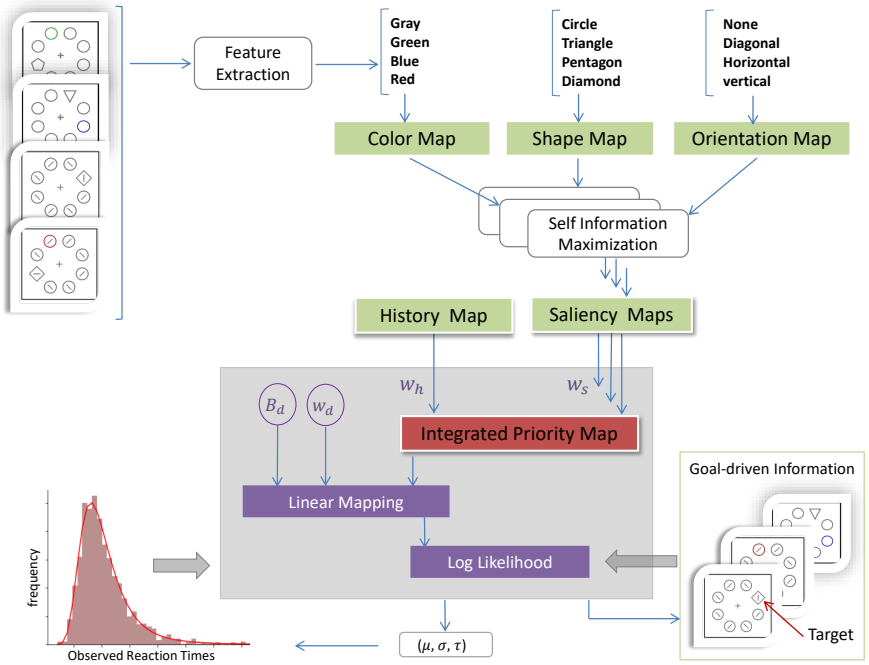

Figure 2: An overview of the algorithmic model. The blue arrows show the direction of data flow from visual input to response and gray arrows show the direction of feedback. w_s, w_h and w_d are map weights. w_s has three elements for color, shape and orientation. w_d has also three elements for distribution parameters(μ, σ, τ). B_d is distribution parameters' bias containing B_μ, B_σ and B_τ.

(such as green or blue). So not only blue, green, triangle and pentagon but also red and diamond are included. In the third version we exclude the history map from the model testing the assumption that only top-down and bottom-up guidance direct attention. To compare these versions of the model, we use a Laplace-approximation. We compute a second-order approximation of the marginal log-probability of the data given the different models' assumptions. We employ these log-probabilites for two Bayesian model comparisons (Bishop, 2006; Barber, 2012; Endres, Chiovetto, & Giese, 2013): fitting one model per participant, and one model per group. In both cases, a model that includes a history map and maps for those features that were predictive during the learning phase is at least 10^{20} as probable as the alternatives. For more details about the model evidences see Figure 4.

Under the assumption that there is a linear mapping from the priority map to the reaction time distribution parameters, the model machine-learns to predict the history map weight (w_h), saliency map weights (w_s) and also the distribution parameters weights and biases (w_d, B_d) (see Figure 2). To compare the weights and also to see how they vary between the color and the shape group see Figure 5, which shows the weights for model version one.

As can be seen in Figure 5, the 'history map' has a higher weight in the color group than in the shape group: to solve the learning task, the color group model has to rely on its learning history features (blue and green) in half of the trials, i.e. in the learning task. Although these colors could be found in the 'color map' as well, there is another color (red) in this map which is task-irrelevant and has to be suppressed. This may be the reason for the increased attention capture by the red distractor in color group members which is reported in (Feldmann-Wüstefeld et al., 2015).

For the search task, a high orientation weight is employed by the color group model, since this task can be solved by spotting an orientation singleton, cf. Figure 1, B.

In contrast, the shape group model can afford to rely less on its 'history map' because the items in its history (triangle and pentagon) exist in the 'shape map' too (triangle, pentagon and diamond), and there is no shape distractor. Therefore, by using a high shape map weight, both the learning task can be solved, and attention can be guided to the shape singleton containing the target in the search task (diamond).

To summarize, the weight of the 'orientation map' is larger in the color group than in the shape group, indicating that the color group model employs orientation saliency in the search task. Using orientation saliency, it does not need to attend to the shape singleton in the search task. However, the shape group model focuses on the 'shape map' which is response-relevant in both tasks.

Also, the weight of the 'color map' was higher in the color group than in the shape group model, since the latter can ig-

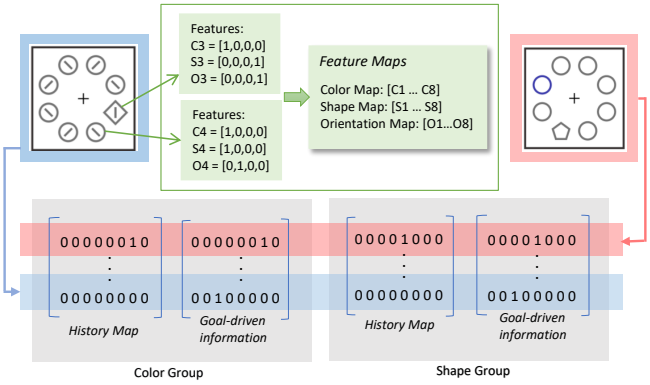

Figure 3: Feature maps, history map and goal-driven information for two random trials. We use 1-out-of-K encoding for the feature vectors, i.e. all components but one are zero. The nonzero component indicates the feature value (see the green box). In each row of history map the location of learned feature is marked. In the target (goal-driven) information the location of response-relevant feature is marked.

nore color altogether.

The model approximates the reaction time distribution parameters (μ, σ, τ) very well (as can be seen in Figure 6). To quantify how close the model-predicted distributions are to the best fit to the data, we evaluate an approximation to the KullbackLeibler (KL) divergence (Bishop, 2006):

$$KL(p||q) = \int p(RT) \log \left(\frac{p(RT)}{q(RT)}\right) dRT \qquad (7)$$
$$\approx \frac{1}{M}\sum_{m=1}^{M} \log p(RT_m) - \frac{1}{M}\sum_{m=1}^{M} \log q(RT_m)$$

where RT_m is the reaction time in trial m, p(RT) and q(RT) are model-predicted and best-fit distributions respectively. For both color and shape group RTs, we find $KL(p||q) \leq 10^{-4}$ which is very close to the minimal possible value.

Conclusion

We presented a model of selection history in visual attention. The model implements the idea that selection history has a role in attention guidance as claimed by Feldmann-Wüstefeld et al. (2015). We compared different versions of the model and the results show that the one which includes selection history, beside bottom-up and top-down control, is best suited for a quantitative description of the behavioral (RT) results. Our model successfully implements an integrated priority map as proposed by Awh et al. (2012). To determine if this integrated priority map approach is indeed the best description of human behavior, future research needs to investigate non-integrated alternatives. Furthermore, as humans use their attention system in a large variety of situations, a model of task switching needs to be added, rather

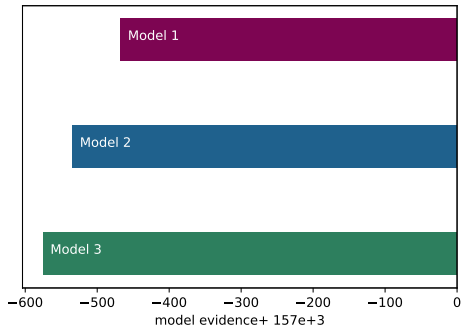

Figure 4: Model comparison. We computed a Laplace-approximation to the Bayesian model evidence across participants. Bigger evidence is better. Model version one, whose history map contains relevant features, scores best. For model descriptions, see text.

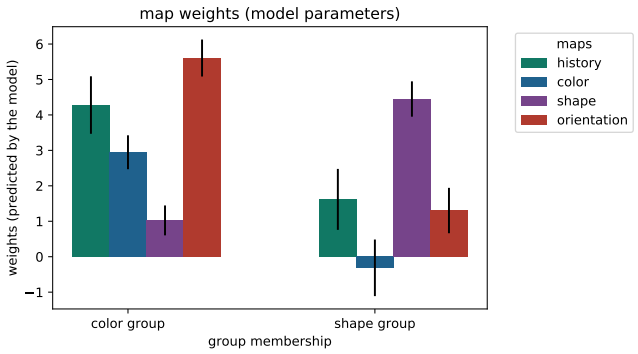

Figure 5: Map weights. For both color group and shape group, optimal map weights for model one are shown. A higher weight means a stronger influence of the corresponding map onto the response and reaction time. The error bars represent the standard deviations of the posterior, i.e. standard errors.

than training one model per task. The search for such alternatives might be facilitated if we knew what the attentional system is actually trying to achieve on a quantitative level. This is a question situated on the 'computational level' (Marr, 1982). Therefore, we intend to build a computational model in a Bayesian/optimal feedback control framework for both ideal and non-ideal observer-actors. Stochastic evidence accumulation approaches – that have been applied in some other models such as Race Models (Mordkoff & Yantis, 1991) and Drift Diffusion models (Luce, 2008) – might be useful to this end. Another interesting avenue of investigation, which would help in constraining the model, would be the addition of physiological variables. For example, adding EEG signals to disentangle target and related sub-processes (such as enhanced target processing or distractor suppression) would shed further light on attentional guidance processes.

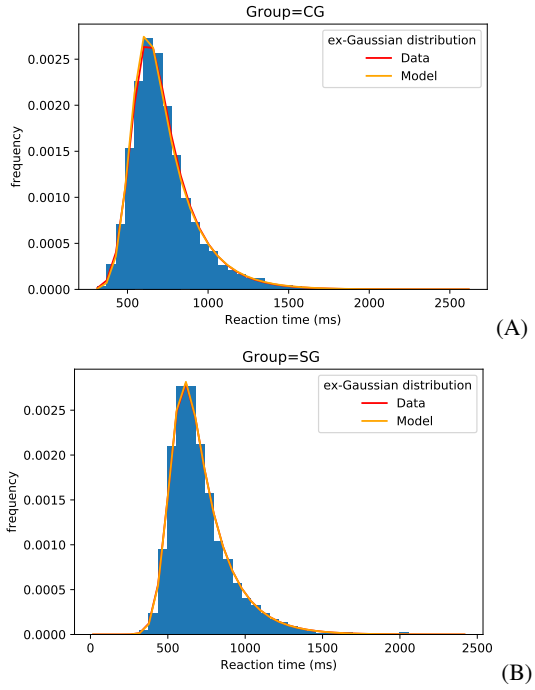

Figure 6: Ex-Gaussian distributions of reaction times. Best fits to the data (red) and model predicted distributions (orange) for participants in (A): color group (CG). (B): shape group (SG).

Acknowledgments

This work was supported by the DFG SFB-TRR 135 "Cardinal Mechanisms of Perception", projects C6 and B3, and "The Adaptive Mind", funded by the Excellence Program of the Hessian Ministry for Science and the Arts. We thank Sara Müller for her contributions to earlier versions of the model.

References

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. Retrieved from http://dx.doi.org/10.1016/j.tics.2012.06.010

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge: Cambridge University Press. Retrieved from http://ebooks.cambridge.org/ref/id/CBO9780511804779

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Borji, A. (2019). *Saliency Prediction in the Deep Learning Era: Successes, Limitations, and Future Challenges.* Retrieved from https://arxiv.org/abs/1810.03716

Borji, A., Sihite, D. N., & Itti, L. (2014). What/Where to Look Next? Modeling Top-down Visual Attention in Complex Interactive Environments. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *44*(5), 523–538. Retrieved from http://ilab.usc.edu/

Bruce, N. D., & Tsotsos, J. K. (2006). Saliency based on information maximization. *Advances in Neural Information Processing Systems*, *18*, 155–162. Retrieved from http://books.nips.cc/papers/files/nips18/NIPS2005_0081.bib

Bruce, N. D., & Tsotsos, J. K. (2009). Saliency , attention , and visual search: An information theoretic approach. *Journal of Vision*, *9*(3), 1–24. Retrieved from https://doi.org/10.1167/9.3.5

Chelazzi, L., Eštočinová, J., Calletti, R., Gerfo, E. L., Sani, I., Libera, C. D., & Santandrea, E. (2014). Altering spatial priority maps via reward-based learning. *Journal of Neuroscience*, *34*(25), 8594–8604. Retrieved from http://doi/10.1523/JNEUROSCI.0277-14.2014

Chikkerur, S., Serre, T., Tan, C., & Poggio, T. (2010). What and where: A Bayesian inference theory of attention. *Vision Research*, *50*(22), 2233–2247. Retrieved from http://dx.doi.org/10.1016/j.visres.2010.05.013

Endres, D. M., Chiovetto, E., & Giese, M. A. (2013). Model selection for the extraction of movement primitives. *Frontiers in Computational Neuroscience*, *7*, 185. Retrieved from https://doi.org/10.3389/fncom.2013.00185

Fecteau, J. H., & Munoz, D. P. (2006). *Salience, relevance, and firing: a priority map for target selection* (Vol. 10) (No. 8). Retrieved from https://doi.org/10.1016/j.tics.2006.06.011

Feldmann-Wüstefeld, T., Uengoer, M., & Schubö, A. (2015). You see what you have learned. Evidence for an interrelation of associative learning and visual selective attention. *Psychophysiology*, *52*(11), 1483–1497. Retrieved from https://doi.org/10.1111/psyp.12514

Henare, D. T., Kadel, H., & Schubö, A. (2020). Voluntary Control of Task Selection Does Not Eliminate the Impact of Selection History on Attention. *Journal of Cognitive Neuroscience*, *32*(11), 2159–2177. Retrieved from https://doi/abs/10.1162/jocn_a_01609

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision*, *9*(5), 1–18. Retrieved from https://doi.org/10.1167/9.5.25

Itti, L., & Borji, A. (2015). *Computational models: Bottom-up and top-down aspects.* Retrieved from http://arxiv.org/abs/1510.07748

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254 – 1259. Retrieved from https://doi.org/10.1109/34.730558

Kadel, H., Feldmann-Wüstefeld, T., & Schubö, A. (2017). Selection history alters attentional filter settings persistently and beyond top-down control. *Psychophysiology*, *54*(5), 736–754. Retrieved from https://doi.org/10.1111/psyp.12830

Kimura, A., Pang, D., Takeuchi, T., Miyazato, K., Yamato, J., & Kashino, K. (2008). *A stochastic model of human visual attention with a dynamic Bayesian network.* Retrieved from http://arxiv.org/abs/1004.0085

Klink, P. C., Jentgens, P., & Lorteije, J. A. M. (2014). Priority Maps Explain the Roles of Value, Attention, and Salience in Goal-Oriented Behavior. *Journal of Neuroscience*, *34*(42), 13867–13869. Retrieved from https://doi.org/10.1523/JNEUROSCI.3249-14.2014

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–227. Retrieved from https://doi.org/10.1007/978-94-009-3833-5

Luce, R. D. (2008). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Company.

Matzke, D., & Wagenmakers, E. J. (2009). Psychological interpretation of the ex-gaussian and shifted wald parameters: A diffusion model analysis. *Psychonomic Bulletin and Review*, *16*(5), 798–817. Retrieved from https://doi.org/10.3758/PBR.16.5.798%0A

Mordkoff, J. T., & Yantis, S. (1991). An Interactive Race Model of Divided Attention. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 520–538. Retrieved from https://doi.org/10.1037//0096-1523.17.2.520

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*(2), 205–231. Retrieved from https://doi.org/10.1016/j.visres.2004.07.042

Ratcliff, R. (1978). A Theory of Memory Retreival. *Psychological Review*, *85*(2), 59–108. Retrieved from https://doi/10.1037/0033-295X.85.2.59

Theeuwes, J. (2018). Visual Selection: Usually Fast and Automatic; Seldom Slow and Volitional. *Journal of Cognition*, *1*(1), 1–15. Retrieved from https://doi.org/10.5334/joc.13

Theeuwes, J. (2019). Goal-driven, stimulus-driven, and history-driven selection. *Current Opinion in Psychology*, *29*, 97–101. Retrieved from https://doi.org/10.1016/j.copsyc.2018.12.024

Todd, R. M., & Manaligod, M. G. (2017). Implicit guidance of attention: The priority state space framework. *Elsevier*, *102*, 121–138. Retrieved from http://dx.doi.org/10.1016/j.cortex.2017.08.001

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. Retrieved from https://doi.org/10.1016/0010-0285(80)90005-5

Tseng, Y. C., Glaser, J. I., Caddigan, E., & Lleras, A. (2014). Modeling the effect of selection history on pop-out visual search. *PLoS ONE*, *9*(3). Retrieved from https://doi.org/10.1371/journal.pone.0089996

Veale, R., Hafed, Z. M., & Yoshida, M. (2017). How is visual salience computed in the brain? Insights from behaviour, neurobiology and modeling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714). Retrieved from https://doi.org/10.1098/rstb.2016.0113

Zelinsky, G. J., & Bisley, J. W. (2015). The what, where, and why of priority maps and their interactions with visual working memory. *Annals of the New York Academy of Sciences*, *1339*(1), 154–164. Retrieved from https://doi.org/10.1111/nyas.12606

APPENDIX C

SECOND MANUSCRIPT

Distracted by Previous Experience: Integrating Selection History, Current Task Demands and Saliency in an Algorithmic Model_

Journal:	*IEEE Transactions on Cognitive and Developmental Systems*
Manuscript ID	TCDS-2022-0555
Manuscript Type:	Regular
Date Submitted by the Author:	01-Dec-2022
Complete List of Authors:	Meibodi, Neda; Psychology Department Abbasi, Hossein Endres, Dominik Schubö, Anna
Keywords:	visual attention modelling, selection histroy, machine learning

Distracted by Previous Experience: Integrating Selection History, Current Task Demands and Saliency in an Algorithmic Model

Neda Meibodi, Hossein Abbasi, Anna Schubö, and Dominik Endres

***Abstract*—Attention can be biased by previous learning and experience. We present an algorithmic-level model of this selection history bias in visual attention that predicts quantitatively how stimulus-driven processes, goal-driven control and selection history compete to control attention. In the model, the output of saliency maps as stimulus-driven guidance interacts with a history map that encodes learning effects and a goal-driven task control to prioritize visual features. We test the model on a reaction-time (RT) data from a psychophysical experiment. The model accurately predicts parameters of reaction time distributions from an integrated priority map that is comprised of an optimal, weighted combination of separate maps. Analysis of the weights confirms selection history effects on attention guidance. The model is able to capture individual differences between participants. Moreover, we demonstrate that a model with a reduced set of maps performs worse, indicating that integrating history, saliency and task information are required for a quantitative description of human attention. Besides, we show that adding intertrial priming effect to the model (as another lingering bias) improves the model's predictive performance.**

***Index Terms*— Ex-Gaussian distribution, Feature integrated theory, Integrated priority map, Selection history, Self information maximization, Visual attention**

I. Introduction And Related Work

SELECTIVE visual attention is a brain function that filters irrelevant sensory inputs to facilitate focusing on relevant items. Stimulus-driven and goal-driven mechanisms have traditionally been proposed to control the process of attention guidance. Object saliency and environment features shape the attentional process in a stimulus-driven manner while the goal-driven process is mostly controlled by observer intentions and preferences.

In addition to goal-driven and stimulus-driven contributions also 'selection history' can play a significant role in guiding attention toward a specific target [1]. Selection history (as a third mechanism of attentional guidance) comes into play when an object is emphasized just because of previous attendance in the same context [2]. To clarify the distinction between goal-driven guidance and selection history, Theeuwes argued that selection history is a fast, effortless, and automatic version of attention control while goal-driven selection is slow and effortful [3]. The term 'selection history' includes several phenomena that can neither be considered as goal-driven nor as stimulus-driven control, such as lingering effects, statistical learning, emotional and also reward-based biases [4].

One special form of selection history has been investigated in [5]–[7]. These studies combined an associative learning task with a visual search task. The results showed that observers attended more to a stimulus they experienced as response-predictive in the preceding feature discrimination task. To examine to what extent selection history can be suppressed by goal-driven process, Kadel *et al.* [6] tested different goal-driven-influenced modes of task preparations such as pretrial task cuing. As their results showed, attentional biases induced by selection history persisted despite task preparation. Their results show that even with these preparations, selection history still plays a noticeable role in biasing attention towards a formerly experienced target. Wolfe and Horowitz [8] mentioned that not only the three aforementioned contributions but also the scene structure and the relative value of the targets and distractors must be considered in modern visual guidance theories. In Guided Search 6.0 [9] –the latest version of the Guided Search model on visual search and selective attention– these five factors are integrated in a spatial priority map to guide attention.

An integrated priority map was also proposed by Awh *et al.* as a theoretical framework for explaining how selection history and other factors of attention guidance interact [1], [2]. Priority map have been successfully employed by many authors [10]–[15] to explain the result of the processes which shape attention. In a review, Klink *et al.* [12] summarized how goal-driven and stimulus-driven maps in cortex combine with a value-based map in midbrain. This combination results in a priority map for the frontal eye fields. Zelinsky and Bisley [11] speculate about the importance of priority map in relationship with visual working memory and also with the motor system. They also higlighted this map as an appropriate construct for predicting behavior.

Stimulus-driven models of attention were developed early on [16]. These models tend to ignore the effects of selection history, task or training [17]. Itti *et al.* [16] implemented feature integration theory (three feature maps including color, intensity and orientation), winner-take-all, inhibition of return

Manuscript submitted December 1, 2022. This work was supported by the DFG SFB-TRR 135 'Cardinal Mechanisms of Perception', project C6 and B3, and 'The Adaptive Mind', funded by the Excellence Program of the Hessian Ministry for Science and the Arts.

Authors are with the Department of Psychology, Philipps-University Marburg, Gutenbergstraße 18, 35032 Marburg, Germany.
E-mails: {neda.meibodi, hossein.abbasi, anna.schuboe, dominik.endres}@uni-marburg.de

and a normalization method to model visual attention in a stimulus-driven manner. This model (or its elaborated version [18]) was subsequently expanded [19], [20]. De Brecht and Saiki [21] showed how Itti and Koch's model [18] can be implemented by neural networks with biologically realistic dynamics based on data from electrophysiology experiments. This model was also expanded later by integrating motion saliency computation [19]. Itti's stimulus-driven model was also combined in a goal-driven model [20] to represent the effect of goal-relevance information on attention or eye-movement. Veale *et al.* [14] validated a neural implementation of Itti's model. In another stimulus-driven model, Bruce and Tsotsos [22] – using self information maximization ($-\log(p(x))$), where x is a feature – proposed a computational model of saliency that is called 'Attention based on Information Maximization (AIM)', because attention is attracted by surprising, i.e. potentially informative, regions of an image.

Most of the models reviewed so far were developed to explain data from highly controlled experiments with impoverished artificial stimuli. However, humans deploy their attention in uncontrolled natural settings replete with complex stimuli. Thanks to deep learning advances, there has been recent progress in deep visual saliency models that can process complex natural images [23]. DeepGaze II is a saliency model that predicts where people look using features from a pre-trained convolutional neural network (VGG-19) and a few layers on top that are trained to read out saliency [24]. While these models have near-human performance compared to observers in front of a screen, they mostly explain saliency effects at their current state of development. It will be interesting to include other attentional guidance mechanisms in them, which go beyond the currently presented scene.

Itti and Borji reviewed more than 50 computational stimulus-driven models [17]. Computational models that integrate goal-driven control [25]–[27] are less well researched than saliency models, likely because they require information not available in the stimulus. Some models integrate stimulus-driven and goal-driven signals in attentional guidance [28]. Chikkerur *et al.* [29] used a Bayesian framework to explain how a combination of stimulus-driven and goal-driven attentional guidance work together in cortex.

Previous studies showed that attention can be biased more toward a target feature which was selected in the last trial ([6], [30]–[32]). This effect, known as intertrial priming, is one of the lingering biases attributed to selection history [3]. Selection history has hardly been modelled despite being a well-known phenomenon. Tseng *et al.* [33] implemented a Ratcliff-type diffusion model [34] for a 2-forced-choice task and showed that intertrial priming can affect diffusion model parameters.

In this paper we introduce an algorithmic-level or 'mechanistic' model (in the sense of Marr [35]) to show how stimulus-driven processes, goal-driven control and selection history compete to guide visual attention toward a specific target[1]. We operationalize selection history as the effect of training-phase learning on the test phase (see [5]–[7]). The model comprises a priority map to integrate goal-driven, saliency-based and history-related biases in a winner-take-all manner. Stimulus-driven guidance, feature maps and saliency maps are made based on feature integration theory [36] and self information maximization [22]. Feature-integration theory, developed by Treisman and Gelade [36], posits that separable dimensions (such as shape and color) are processed separately before being integrated on-demand. Using this theory, the proposed model, codes the input into three types of features (color, shape and orientation) and computes a saliency map for each feature dimension with AIM [22]. The model also incorporates the effect of intertrial priming which emphasises the response-relevant feature dimension of the last trial in the current one [37]. Additionally, a history map contributes to the integrated priority map to reflect the effect of selection history and learning in the model. Finally, task-relevant information controls the map integration weights that generate predictions for responses and response times. These integration weights are our model for the goal-driven influences. We test this model on a behavioral database from an experiment by Feldmann-Wüstefeld *et al.* [5]. The model can predict the reaction time distribution parameters for each participant and also across the experimental groups. To find the best fitted distribution on reaction times, several probability density functions are compared minimizing negative log likelihood and the best fitting one –an ex-Gaussian distribution [38]– is used in the model.

The rest of the paper is organized as follows: we review the experiment and explain its details required for a full understanding of the model. We then compare models that differ in the information that enters into the integrated priority map and show that a model with selection history information –on feature level– performs best. We also show that the inclusion of intertrial priming variables lead to an increase of the (approximate) Bayesian model evidence. More information about data analysis and reaction time distributions can be found in Appendix A and B.

II. Experimental Data

The data used in this study comes from the first experiment of Feldmann-Wüstefeld *et al.* [5]. They investigated the impact of associative learning on covert selective visual attention to examine whether selection history effects generalize from particular features (e.g., 'blue' or 'green') to the entire color dimension. The experiment consisted of a 'practice' and a 'main' phase, in which two types of tasks (learning and search) were performed. A central fixation cross was presented on the screen, which was then surrounded by eight different elements on an imaginary circle (Fig. 1). 28 participants were divided randomly into two different groups, namely 'color group' and 'shape group'. They were first naive about their group membership, but had to learn it on a trial and error basis in the practice phase.

In the 'practice phase', participants had to learn that either color or shape was the response-relevant dimension in this learning task (see Fig. 1.a). Members of the color group had to report the color of the color singleton (blue or green), whereas

[1]A preliminary version of this modeling study has been presented at COGSCI 2021 https://cognitivesciencesociety.org/cogsci-2021/.

members of the shape group had to respond to the shape of the shape singleton (triangle or pentagon). They had to use their left hand to press one of two buttons that were placed on the left side of the response pad. Auditory feedback indicated whether they pressed the incorrect key.

In the 'main phase' a second visual search task was added, and participants performed both tasks in random order. In the search task (Fig. 1.b), all participants had to search for a shape target and report the orientation of a line presented inside the diamond shaped target. In half of the trials, a response-irrelevant red circle was presented as distractor. Participants used their right hand to press one of two buttons on the right side of the pad to indicate the line orientation (horizontal versus vertical).

The results of this study showed that the history of selection acquired in the learning task affected the participants' performance in the search task. Reaction time analysis showed that responses in distractor-present trials were slower than in distractor-absent trials, and the distractor cost was larger in color group participants than in participants in the shape group. Concurrently recorded EEG signals also suggested that participants in the color group deployed attention towards the red color distractor, this was not the case for participants in the shape group. Accordingly, the authors suggested that the participants' history of either shape or color selection in the practice phase had resulted in a selection history bias. Feldmann-Wüstefeld *et al.* had done their study in 4 experiments to examine the influence of task switching. In the first and the second experiments, learning and search trials were intermixed. In experiment three, the tasks were presented block-wise and in the forth one, the tasks were performed on separate days. The results of all experiments demonstrated the presence of selection history effect on attention deployment even when the tasks were done on different days [5]. We decided to use the intermixed presentation trials to model the effect of repeating the response-relevant feature dimension from trial n-1 to trial n along with goal-driven, stimulus-driven and selection history influences on attentional control.

We developed a model of this selection history bias in the current study based on the behavioral data from the main phase, which comprises at total of 28672 trials across all participants. More details about the experiment can be found in [5].

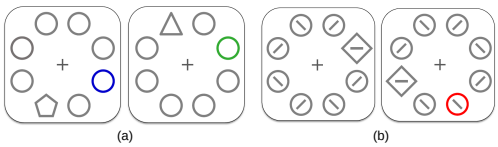

Fig. 1. Experiment displays. Learning task (a): Participants in the color group had to respond to the color (green vs. blue) and participants in the shape group had to respond to the shape (pentagon vs. triangle). Search task (b): The orientation (horizontal vs. vertical) of the line embedded in the diamond had to be reported in distractor-absent (left) and distractor-present trial (right).

III. The Algorithmic Model

We first determined which trial parameters affect participants' RTs by reanalyzing the data from [5]. This served two purposes: replicated the reported selection history effects with a different analysis method, and determined relevant variables for the more detailed model presented later on. The results of this analysis showed that the group membership – which we modeled as a history map– and also trial types – which were coded into saliency maps– both influence RTs. By 'trial type' we mean if the trial is either a learning or a search (distractor-absent, distractor-present) type. For more information about this analysis, see Appendix A. The analysis also showed that switching from one task type to another can affect participants' RTs and that this effect differs between experimental groups (see Fig. 10). Since the model employs the ex-Gaussian as an RT distribution, we detail in Appendix B how this distribution was chosen based on a model comparison.

We assembled an algorithmic-level model to explain how goal-driven and stimulus-driven influences competitively interact with visual selection history to guide attention toward a specific stimulus. Inspired by the integrated priority map in [2], we used a 'history map' reflecting the influence of selection history on current attention deployment, see Fig. 2. Additionally, there is an overall saliency map for stimulus-driven influences. How these maps combine into an integrated priority map is determined by task-dependent weights. Fig. 2 also shows how the output of the integrated priority map is used to predict ex-Gaussian distribution [39] parameters of reaction times (left exit path in the figure) and response likelihoods (the right exit path). Evaluating these response likelihoods and reaction times against participants' reaction times allows us to fit the model to the experimental data, see (5) below.

The input stage of the model is based on assumptions made by visual search theories such as feature-integration [36] and guided search [9]. The model extracts three types of features (color, shape and orientation) and feature maps –as shown in Fig. 2– are computed. In the next processing step, saliency maps that model the effect of stimulus-driven control on visual attention [40] are formed from the feature maps. Shannon's measure of Self-Information is applied, similar to Attention Based on Information Maximization [22], to compute saliency maps. Equations (1) and (2) show the actual calculations behind map computation. Feature maps are $M \times N \times K$ vectors where M is the number of trials, N is the number of objects in each trial and K is the number of distinct values that each feature can take on, i.e. we are using 1-out-of-K encoding for the features, with the value 1 indicating which feature value is present. In the current experiment $M = 1024$ (for each participants), $N = 8$ and $K = 4$. Fig. 3 illustrates the method of building feature maps for some example trials. For all trials, we take the feature maps f_i for $i \in \{color, shape, orientation\}$ and compute the self-information X_i:

$$\forall k : X_i[k] = -log\left(\sum_{n=1}^{N} f_i[n][k]/N\right) \tag{1}$$

which yields the saliency of all trials $s_i[n]$:

$$\forall n : s_i[n] = X_i\left[\arg\max_k(f_i[n][k])\right] \tag{2}$$

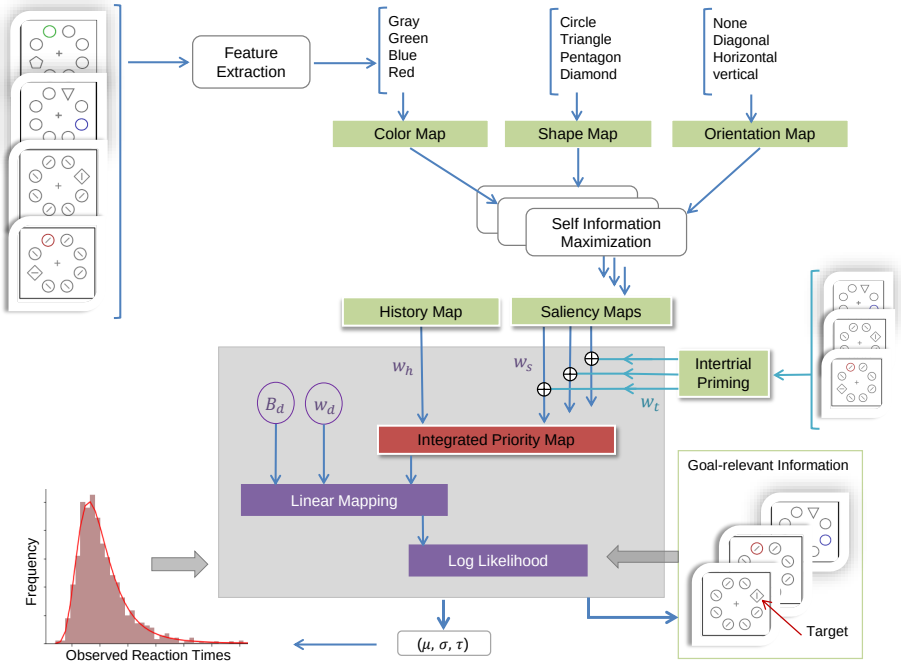

Fig. 2. An overview of the algorithmic model. The blue arrows show the direction of data flow from visual input to response and gray arrows show the direction of feedback.w_s, w_h, w_t and w_d are model parameters that weigh how strongly different maps enter into the integrated priority map or the RT prediction. w_s is saliency weight with three elements for color, shape and orientation. w_t is intertrial priming weight and also has three elements for color, shape and orientation. w_h is history map weight. w_d is distribution parameters weight and has three elements for μ, σ and τ. B_d is distribution parameters' bias containing B_μ, B_σ and B_τ. The goal-relevant information (on the right side of the figure) helps the model to guide attention to the target location.

where, due to the 1-of-K feature encoding, we can use $argmax$ to pick the self-information corresponding to the current feature value.

Saliency maps s_i are fed into the integrated priority map along with history information (h) to compete in a soft winner-take-all model [1] for the predicted response target. Selection history, the third category of attentional guidance [2], carries the effect of learning (participants learned about color or shape in our experiment) into the priority map (p). To model the priming effect of the last trial on participant' RT, we added another parameter (w_t) to the model. The parameter w_t includes a weight for each feature dimension and it modulates the saliency weights when the maps are combined into the priority map:

$$\forall m, n : p[m][n] = \underset{n}{\text{soft}\max}\left(\sum_i \left((w_{s_i} + w_{t_i} * t_i[m]) * s_i[m][n]\right) + w_h * h[m][n]\right) \quad (3)$$

The weights (w_h for history and w_{s_i} for $i \in \{color, shape, orientation\}$) are used to combine the history map and the saliency maps computed from from color, shape and orientation. These weights reflect the influence of the content of the respective map on the integrated priority map for the tasks that the model will be optimized for. In (3) t is a $M \times i$ matrix which carries information from the last trial: in each row of t, a '1' indicates the feature dimension which had to be selected by the participants in the last trial (see Fig. 4). The softmax function is used to ensure that the winning location receives the most attention while keeping the map interpretable as a probability distribution.

In our model, (3) can be interpreted as the first layer of a (two-layer) neural network. The second layer is a (linear) mapping from the integrated priority map to reaction time distribution parameters:

$$\forall m : d[m] = \sum_{n=1}^{N} (p[m][n] * w_d) + B_d \quad (4)$$

Where w and B are weights and biases of ex-Gaussian distribution parameters' $d[m] \in (\mu[m], \sigma[m], \tau[m])$ for each trial m.

We also compute a 1-out-of-K representation of the goal-relevant information (g in (5)) (see Fig. 3) which is used for machine-learning the weights with which the history map and the saliency maps are combined in the priority map. Psychologically, we can interpret the role of this as combined guidance of stimulus-driven and history toward the location of the target.

The weights (w_h, w_s, w_t and w_d) for both tasks are determined by maximizing the log of the joint distribution of the reaction times (RT), the goal-relevant information (g) under the distribution predicted by the integrated priority map (p) and eventually the prior distributions over the model

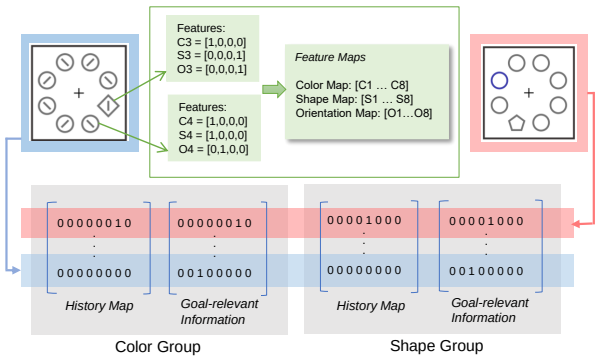

Fig. 3. Feature maps, history map and goal-relevant information for two random trials. We use 1-out-of-K encoding for the feature vectors, i.e. all components but one are zero. The nonzero component indicates the feature value (see the green box). In each row of history map the location of learned feature is marked. In goal-relevant information the location of response-relevant feature is marked.

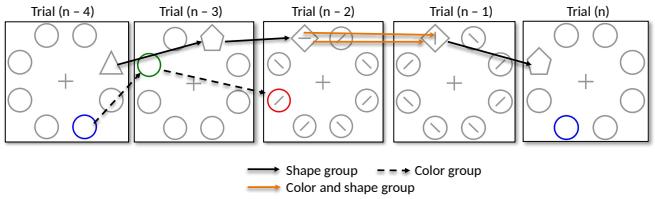

Fig. 4. Intertrial priming. The priming effect in our model is depicted by solid lines for shape, and dashed lines for color. Attending a particular feature (e.g. triangle shape in trial n-4) primes shape dimension attention in the next trial, here: pentagon in trial n-3. Likewise for color dimension. Line orientation priming is possible too, shown for trial n-2 → trial n-1 (upper orange arrow). Note that shape priming is possible in the color group too if two search trials follow each other (lower orange arrow).

parameters δ:

$$\begin{aligned} L &= \sum_{m=1}^{M} \log\left(ExG\left(RT[m] \mid \mu[m], \sigma[m], \tau[m]\right)\right) \\ &+ \sum_{m=1}^{M}\sum_{n=1}^{N} \left(\log p[m][n] * g[m][n]\right) + \delta \end{aligned}$$

where ExG is the ex-Gaussian density function. δ is computed as the sum of the logs of the following prior distributions on the parameters:

$$\begin{aligned} w &\sim \mathcal{N}(0.0, 1.0) \\ B_{\mu} &\sim \mathcal{N}(600.0, 100.0) \\ B_{\sigma^2} &\sim \mathcal{N}(75.0, 4.0) \\ B_{\tau} &\sim \mathcal{N}(200.0, 20.0) \end{aligned} \tag{5}$$

Mean and standard deviation of the last three distributions are selected in a way that matches results from similar experiments [5], [6]. To find the weights and biases that maximize the joint probability (5), we draw random initial values from these distributions and then optimize using Python 3.8.8, PyTorch 1.8.1 and Adam optimizer with learning rate 0.1.

IV. Results and Discussion

To investigate how selection history and saliency maps quantitatively predict attentional guidance, we tested seven versions of the model. In the first model (M1 in Fig. 5), the history map contains the response-relevant features of the learning phase (blue and green for the color group, triangle and pentagon for the shape group). This model is used as the basis for models M2 to M7, which are altered versions thereof. In M2 the history map includes all color singletons (for participants in the color group) and all shape singletons (for participants in the shape group). Here, the assumption is that the participants have learned response-predictiveness on the dimensional level (color or shape), not on the level of single features (such as green or blue). So not only blue, green, triangle and pentagon but also red and diamond are included. In M3, priming information from previous trials is removed from the model. In M4, we exclude the history map from the model testing the assumption that only goal-driven and stimulus-driven guidance direct attention. In M5, M6 and M7, shape, color and orientation maps are removed to see if all saliency maps are needed to explain the experimental data. To compare these versions of the model, we use a Laplace-approximation to the Bayesian model evidence. We compute a second-order approximation of the marginal log-probability of the data given the different models' assumptions. We employ these log-probabilities for two sets of model comparisons [41]–[43]: fitting one model per participant, and one model per group. In both cases, comparison shows M1 being the most probable, which includes saliency maps, a history map with the features that were predictable during learning and also the effect of the last trial on the current one (as the intertrial priming effect). This model is at least 10^{20} times more probable than the alternatives. For more details about the model evidences see Fig. 5.

Under the assumption that there is at least an approximately linear mapping from the priority map to the reaction time distribution parameters, the model machine-learns to predict the history map weight (w_h), saliency map weights (w_s), intertrial priming weights (w_t) and also the distribution parameters weights and biases (w_d, B_d) (see Fig. 2). A comparison of the learned weights and their differences between the color group and the shape group is shown in Fig. 6. As one might expect, the color weight is higher in the color group, whereas the shape weight dominates in the shape group. This leads to a stronger influence of the respective saliency map on the contents of the integrated priority map, which is shown in Fig. 7 for a distractor-present trial. In other words, while we assume that saliency is a property of the physical stimulus statistics, the weight with which saliency enters into the integrated priority map can be varied by (learnable) task demands. In Fig. 7 the individual map acivations and their weighted combinations are shown in color coding. The color group model's attention is strongly drawn towards the (red) color distractor. In contrast, the shape group model prioritizes the correct target location.

As Fig. 6 shows, the 'history map' has a higher weight (w_h) in the color group than in the shape group: to solve the learning task, the color group model has to rely on the colors

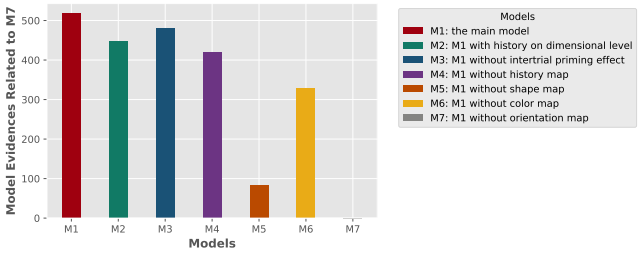

Fig. 5. Model comparison. We computed a Laplace-approximation to the Bayesian model evidence across all participants. The evidences are plotted relative to M7 (the least probable model). Bigger evidence is better. M1, which is called the main model, contains the saliency maps, intertrial priming effect and the history map on response-relevant features. This model scores best. For models descriptions, see the text.

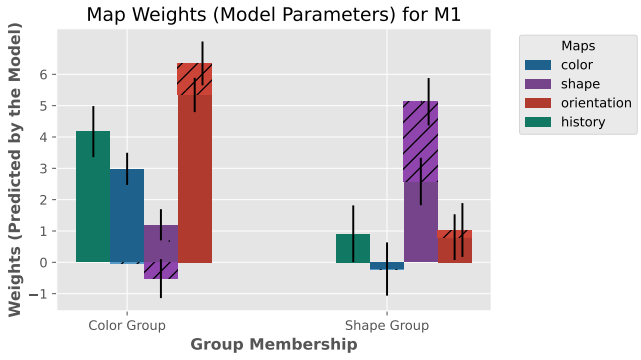

Fig. 6. Map weights. For both color group and shape group, optimal map weights for the first model are shown. A higher weight means a stronger influence of the corresponding map onto the response and reaction time. The hatched parts are the weights modulations by the intertrial priming. Note that the final weight of each saliency map is a sum over the map weight and the priming weight, see (3) and also Fig. 2. Priming modulations on color maps are very small (close to zero) and can be hardly seen. Priming modulation on shape map in color group is negative. The error bars represent the standard deviations of the posterior, i.e. standard errors.

(blue and green) encountered during the practice phase which is reflected in the large weight of the history map. Although these colors could be found in the 'color map' as well, there is another color (red) in this map which is task-irrelevant and has to be suppressed, hence the smaller weight of the color map. This is the reason for the increased attention capture by the red distractor in color group members which is reported in [5]. In other words, the presence of a color distractor leads to a down-weighting of dimensional color saliency in favor of a feature-level color representation. For the search task, a high orientation weight is employed by the color group model, since this task can be solved by spotting an orientation singleton.

In contrast, the shape group model can afford to rely mostly on the 'shape map', because the items in its history (triangle and pentagon) exist in the 'shape map' too (triangle, pentagon and diamond), and there is no shape distractor. Therefore, by using a high shape map weight, both the learning task can be solved and attention can be guided to the shape singleton containing the target in the search task (diamond).

To summarize, the weight of the 'orientation map' is larger in the color group than in the shape group, indicating that the color group model relies on orientation saliency in the search task. However, the shape group model focuses on the 'shape map' which is response-relevant in both tasks. Also, the weight of the 'color map' was higher in the color group than in the shape group model, since the latter group can ignore color altogether.

All versions of the model except M3 can capture intertrial priming. To see how intertrial priming effect is defined in our model please see Fig.4. Our assumption is that intertrial priming is dimension-specific rather than feature-specific. This is also claimed by Liesefeld *et al.* [44]. In our model, intertrial priming has three weights (w_t) for color, shape and orientation priming. See also (3) and the hatched parts in Fig.6. In the color group the modulatory effect of intertrial priming causes an increased orientation map weight and also a reduction of the shape map weight. The former indicates that the generally high reliance of a color group model on orientation during a search task is amplified during repetition of search task trials. The latter might represent task switching: switching from the search task (reporting orientation embedded in a shape singleton target) to the learning task (reporting colors and not shape singletons) is best accomplished by down-weighting shape features temporarily. Interestingly, for color group participants, there is no priming-driven weight modulation of color map. Irrespective of the previous trial's type, a color group model relies more on the history map than on the color map to ignore the red distractor. In our opinion, this rules out the alternative hypothesis that longer response times in the color group are induced by task switching efforts only, and not by selection history and the need to suppress the red distractor. This is in agreement with the results of experiments 3 and 4 reported in [5]. In both of these task variants, learning and search task were separated, either block-wise, or by asking participants to perform the tasks on separate days. Search performance of the color group, however, was still affected by their prior selections in the learning task, even though participants now performed only search tasks trials, and task switching no longer occurred.

The model approximates the reaction time distribution parameters (μ, σ, τ) very well (as can be seen in Fig. 8). To quantify how close the model-predicted distributions are to the best fit to the data, we evaluate an approximation to the Kullback–Leibler (KL) divergence [41]:

$$KL(p||q) = \int p(RT) \log \left(\frac{p(RT)}{q(RT)}\right) dRT \qquad (6)$$
$$\approx \frac{1}{M}\sum_{m=1}^{M} \log p(RT_m) - \frac{1}{M}\sum_{m=1}^{M} \log q(RT_m)$$

where RT_m is the reaction time in trial m, p(RT) and q(RT) are model-predicted and best-fit distributions respectively. For both color and shape group RTs, we find $KL(p||q) \leq 10^{-4}$ which is very close to the minimal possible value of zero.

V. Conclusion

The presented model shows how saliency, selection history and goal-driven demands collaborate in guiding visual attention. The model implements the idea that selection history

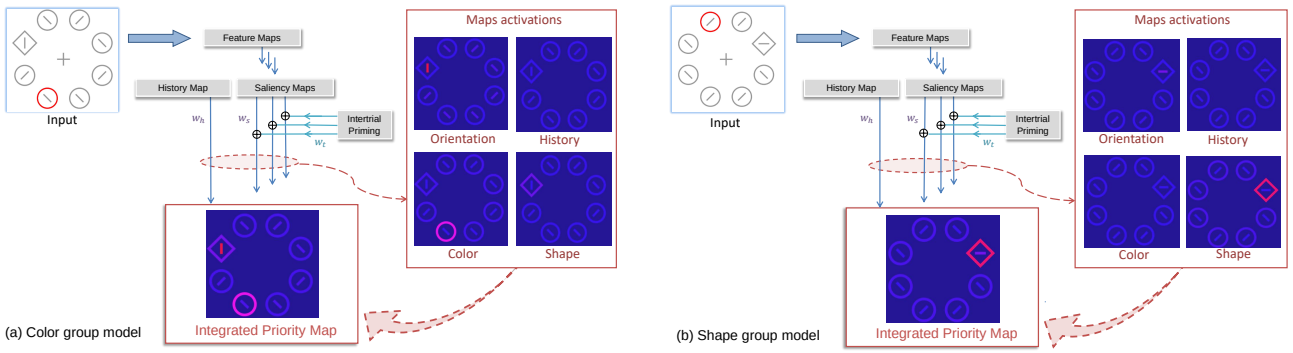

Fig. 7. Map activations in distractor-present search task trials for a color group (a) and a shape group (b) participant model. To visualize these activations the weighted value of each map is used as a color code: $((w_{s_i} + w_{t_i}) * s_i)$ for each saliency map (i) and $(w_h * h)$ for history map. See Fig. 2 for variable names. Warmer colors indicate higher activations. Individual weighted map activations are integrated in the final priority map and attention is guided to the location with the highest feature activity. See (3) for a computational description.

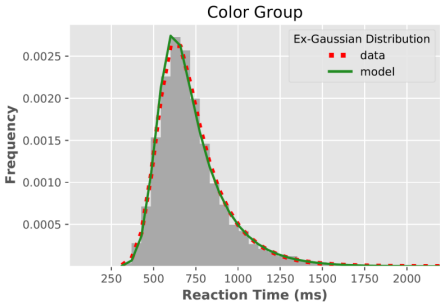

Fig. 8. Ex-Gaussian distributions of reaction times. Best fits to the data (red) and model predicted distributions (green) for participants in the color group. Shape group model predictions fit comparably well.

plays an important role in attention guidance as claimed in [5]. We compared different versions of the model and the results show that the one which includes selection history (long–learned selection preferences and also intertrial priming), beside stimulus-driven and goal-driven control, is best suited for a quantitative description of the behavioral (RT) results. This paper is our first effort to model selection history as an attentional mechanism. To make the model more comprehensive, we plan the following future steps: since previous experiments on selection history effects were done with impoverished stimuli and simple participant responses, we planned to run an experiment in natural or semi-natural (virtual reality) environments. The data obtained in richer environments will likely require an extension of the model, in particular with respect to stimulus representation and response capability. Second, our model successfully implements an integrated priority map [2], [45]. To determine if this integrated priority map approach is indeed the best description of human behaviour, future research needs to investigate non-integrated alternatives. The search for such alternatives might be facilitated if we knew what the attentional system is actually trying to achieve on a quantitative level. This is a question situated on the 'computational level' [35]. Therefore, we intend to build a computational model in a Bayesian/optimal feedback control framework for both ideal and non-ideal observer-actors. Stochastic evidence accumulation approaches – that have been applied in some other models such as Race Models [46] and Drift Diffusion models [39] – might be useful to this end. Besides, to have dynamical version of the model, we are building a dynamical priority map which is updating over time. Importantly, our model does not yet include an explicit reinforcement learning component. Participants did learn the tasks from negative reinforcement only. Our model captures participants' behaviour after this reinforcement learning phase is completed, which was determined by a high enough performance level (see [5]). We agree that it would be interesting to model this first phase in future work, too. Another interesting avenue of investigation, which would help in constraining the model, would be the addition of physiological variables. For example, adding EEG signals to disentangle processes of target selection and distractor suppression would shed further light on attentional guidance processes.

Appendix A
Statistical Analysis of Reaction Times

We analyzed reaction times (RT) binarized at their grand (population) mean across all trial types and participants. While this is a fairly coarse analysis, it should reveal strong effects of experimental manipulations. Thus, we model the probabilistic dependence of these binarized times on different variables with Bayesian networks. The variables we consider are group membership, trial type or both. By group membership we mean either 'color group (CG)' or 'shape group (SG)', i.e. whether participants had to categorize either color or shape singletons in the practice phase of the experiment.

Evaluating a network with both trial type and group membership as conditioners of reaction time, one can see the joint effect, (see Fig. 9.a and 9.b). Although both groups are slower in distractor-present trials, color group participants are less probable to have reaction times smaller than the grand mean. This may be a selection history effect: these participants learned to respond to color, and they are consequently more distracted by a color distractor.

To see how 'intertrial priming effect' is reflected in behavioral data, we conducted another analysis (see Fig. 10). Here, we wanted to know if switching from one trial to another

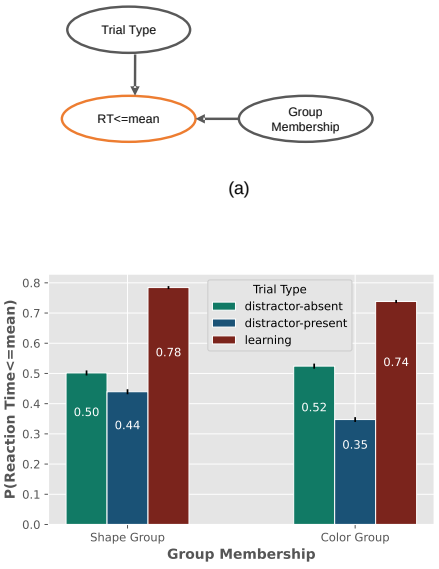

Fig. 9. Reaction time analysis. The Bayesian network shows the tested dependencies between experimental variables (group membership and trial types) and reaction times. The bar chart shows the probability of having a reaction time less than the grand mean. Error bars represent the standard error of the mean. Trial type has a strong effect on reaction time. Color group participants are more distracted by a color singleton than shape group members.

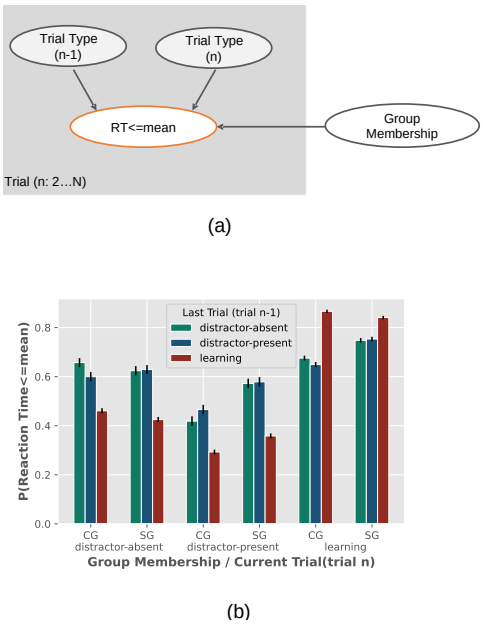

Fig. 10. Reaction time analysis after adding intertrial priming variables : a) the network models dependencies between RT, trial types and group membership. N is the number of trials, trial n is the current trial and trial n-1 is the previous trial. b) the probability of having RT less than or equal to the grand mean is shown on the bar-charts based on the group membership, current trial type and last trial type. RT probability reduction (mean RT increase) is particularly strong for color group members when a learning trial is followed by a distractor-present trial, possibly because their selection history is activated. Error bars represent the standard error of the mean.

impacts the probability of having RT less than mean and additionally if this impact differs between groups. As the bar charts in (Fig. 10.b) show, participants are faster when the trial type is duplicated (i.e., the current trial (trial n) is the same as the last one (trial n-1)). Switching between learning and search trials always causes delays in responding and it is especially noticeable in distractor-present trials when the last trial is a learning trial. Considering the group membership as another RT conditioner, we see that the above mentioned observation is more pronounced in color group (see bar charts in Fig. 10.b):

$$P(RT \leq mean |\ T_n = distr.\ pres.,\ T_{n-1} = learn., \\ |\ group\ memb. = CG) = 0.29 \quad (7)$$

where N is the number of trials in these condition and T is trial type. This probability is 0.35 for shape group. These reduced probabilities might be due to attending the distractor in both groups. However, the reduction is stronger in the color group, where the (color) selection history is activated by the previous (T_{n-1}) learning trial. In other words, responding to the color in a learning trial increases the chance of being distracted by the red distractor in the subsequent search trial, and thus responding to a shape singleton takes longer.

Appendix B
Reaction Time Distribution

Reaction time measurements have been widely employed in psychological experiments to analyze behavioral responses to well-defined tasks. Psychologists agree that there are three main types of reaction times: simple reaction times, recognition reaction times, choice reaction times and also some more forms such as discrimination reaction times and decision reaction times that come from combining varieties of experimental tasks [47].

Many distributions have been used to describe RT in neurocognitive and psychological research. In [48], the Gamma distribution is used to model PEBL (Psychology Experiment Building Language) Go/No-Go tests, with the primary motivation that RTs can be modeled better with a right-skewed distribution. In another study, inverse Gaussian (Wald) is used in a theoretical analysis of psychophysical parameters in a 2AFC design [49] with the assumption that if RT is the time needed for an evidence accumulation to reach a fixed boundary –similar to Brownian diffusion process– it is distributed as an inverse Gaussian. Another popular distribution is the Recinormal [50] which is introduced in LATER model to describe psychological decision making processes [51], [52].

We tried to find the best fitting distribution model for the reaction times in our data by testing the following distribution types against each other by approximate Bayesian model comparison: Gaussian, Gamma, inverse Gamma, inverse Gaussian, Recinormal and exponential Gaussian(ex-Gaussian). We fitted the distributions on each participant's RT and also on aggregates defined by group memberships and trial types. We then compared the negative log likelihoods, which showed that the best fit is achieved with an ex-Gaussian. For illustration, Fig. 11 shows the different densities (red lines) fitted to one

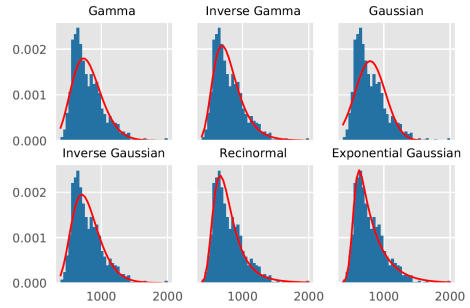

Fig. 11. Distribution functions fitted on RT datasets for a random participant. Ex-Gaussian fits best.

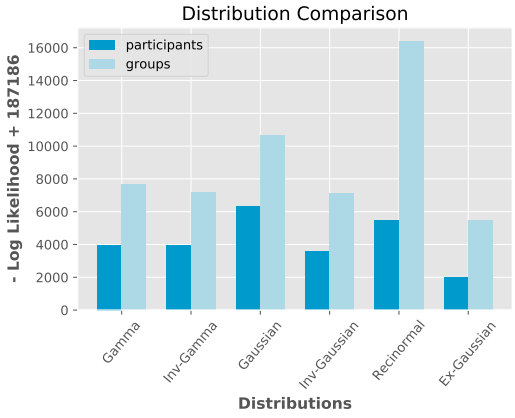

Fig. 12. Distribution comparison using -log likelihood score, lower is better. Distributions are fitted on individual participants data and on group-based data (all color group participants and all shape group participants). The best distribution is Ex-Gaussian, followed by inverse Gaussian.

of the participant's RT data (histograms) . Visual inspection indicates that the ex-Gaussian provides the best fit. For a quantitative comparison we computed each distribution's parameters that minimized the sum of negative log likelihood scores (participants-based and also group-based). These scores are depicted in Fig. 12 , lower is better. The best fits are ex-Gaussian and inverse Gaussian models respectively. To fit the distributions on data we used Python 3.8.8 and PyTorch 1.8.1.

The ex-Gaussian distribution is a convolution of Gaussian and exponential distributions, see e.g. [39]. It has three parameters: μ, σ and τ that are the mean and standard deviation of the Gaussian component and the mean of the exponential component, respectively. The mean and the variance of this distribution are $\mu+\tau$ and $\sigma^2 + \tau^2$.

Response times are not distributed normally [53]. Because of their long tail on the right, RT distributions might have an exponential component. Christie and Luce [54] and also McGill [55] therefore proposed that RT distributions are a convolution of two components. For an ex-Gaussian distribution to arise, one of them has to have an exponential distribution. The above mentioned authors had opposite believes about the source of this exponential component: Christie and Luce mentioned that decision time is exponentially distributed but McGill related that to movement response. Hohle [56] also tried to show that the RT distribution is a convolution of a normal and exponential components by auditory RT experiments.

The ex-Gaussian has been popular recently in psychological research. In [57] several distributions are fitted on behavioral data of three visual search tasks. The best fits all have an exponential component and the ex-Gaussian is one of those. Ex-Gaussian parameters can even be useful in evaluating attention disorders [58], [59]. More research on ex-Gaussian parameters analysis can be found in [60], [61].

REFERENCES

[1] J. Theeuwes, "Goal-driven, stimulus-driven, and history-driven selection," *Current Opinion in Psychology*, vol. 29, pp. 97–101, 2019. doi: 10.1016/j.copsyc.2018.12.024

[2] E. Awh, A. V. Belopolsky, and J. Theeuwes, "Top-down versus bottom-up attentional control: a failed theoretical dichotomy," *Trends in Cognitive Sciences*, vol. 16, no. 8, pp. 437–443, 2012. doi: 10.1016/j.tics.2012.06.010

[3] J. Theeuwes, "Visual Selection: Usually Fast and Automatic; Seldom Slow and Volitional," *Journal of Cognition*, vol. 1, no. 1, pp. 1–15, 2018. doi: 10.5334/joc.13

[4] M. Failing and J. Theeuwes, "Selection history: How reward modulates selectivity of visual attention," *Psychonomic Bulletin and Review*, vol. 25, no. 2, pp. 514–538, 2018. doi: 10.3758/s13423-017-1380-y

[5] T. Feldmann-Wüstefeld, M. Uengoer, and A. Schubö, "You see what you have learned. Evidence for an interrelation of associative learning and visual selective attention," *Psychophysiology*, vol. 52, no. 11, pp. 1483–1497, 2015. doi: 10.1111/psyp.12514

[6] H. Kadel, T. Feldmann-Wüstefeld, and A. Schubö, "Selection history alters attentional filter settings persistently and beyond top-down control," *Psychophysiology*, vol. 54, no. 5, pp. 736–754, 2017. doi: 10.1111/psyp.12830

[7] D. T. Henare, H. Kadel, and A. Schubö, "Voluntary control of task selection does not eliminate the impact of selection history on attention," *Journal of Cognitive Neuroscience*, vol. 32, no. 11, pp. 2159–2177, 2020. doi: 10.1162/jocn_a_01609

[8] J. M. Wolfe and T. S. Horowitz, "Five factors that guide attention in visual search," *Nature Human Behaviour*, vol. 1, no. 3, pp. 1–8, 2017. doi: 10.1038/s41562-017-0058. [Online]. Available: http://dx.doi.org/10.1038/s41562-017-0058

[9] J. M. Wolfe, *Guided Search 6.0: An updated model of visual search*. Psychonomic Bulletin & Review, 2021, vol. 28, no. 4. ISBN 1342302001859

[10] J. H. Fecteau and D. P. Munoz, "Salience, relevance, and firing: a priority map for target selection," *Trends in Cognitive Sciences*, vol. 10, no. 8, pp. 382–390, 2006. doi: 10.1016/j.tics.2006.06.011

[11] G. J. Zelinsky and J. W. Bisley, "The what, where, and why of priority maps and their interactions with visual working memory," *Annals of the New York Academy of Sciences*, vol. 1339, no. 1, pp. 154–164, 2015. doi: 10.1111/nyas.12606

[12] P. C. Klink, P. Jentgens, and J. A. M. Lorteije, "Priority maps explain the roles of value, attention, and salience in goal-oriented behavior," *Journal of Neuroscience*, vol. 34, no. 42, pp. 13 867–13 869, 2014. doi: 10.1523/jneurosci.3249-14.2014

[13] R. M. Todd and M. G. Manaligod, "Implicit guidance of attention: The priority state space framework," *Elsevier*, vol. 102, pp. 121–138, 2017. doi: 10.1016/j.cortex.2017.08.001

[14] R. Veale, Z. M. Hafed, and M. Yoshida, "How is visual salience computed in the brain? Insights from behaviour, neurobiology and modeling," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 372, no. 1714, 2017. doi: 10.1098/rstb.2016.0113

[15] L. Chelazzi, J. Eštočinová, R. Calletti, E. L. Gerfo, I. Sani, C. D. Libera, and E. Santandrea, "Altering spatial priority maps via reward-based learning," *Journal of Neuroscience*, vol. 34, no. 25, pp. 8594–8604, 2014. doi: 10.1523/JNEUROSCI.0277-14.2014

[16] L. Itti, C. Koch, and E. Niebur, "A model of saliancy-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254 – 1259, 1998. doi: 10.1109/34.730558

[17] L. Itti and A. Borji, "Computational models: Bottom-up and top-down aspects," 2015. [Online]. Available: http://arxiv.org/abs/1510.07748

[18] L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vision Research*, vol. 40, no. 10-12, pp. 1489–1506, jun 2000. doi: 10.1016/S0042-6989(99)00163-7

[19] D. F. Ramirez-Moreno, O. Schwartz, and J. F. Ramirez-Villegas, "A saliency-based bottom-up visual attention model for dynamic scenes analysis," *Biological Cybernetics*, vol. 107, no. 2, pp. 141–160, 2013. doi: 10.1007/s00422-012-0542-2

[20] J. Tanner and L. Itti, "A top-down saliency model with goal relevance," *Journal of Vision*, vol. 19, no. 1, p. 11, 2019. doi: 10.1167/19.1.11

[21] M. de Brecht and J. Saiki, "A neural network implementation of a saliency map model," *Neural Networks*, vol. 19, no. 10, pp. 1467–1474, 2006. doi: 10.1016/j.neunet.2005.12.004

[22] N. D. B. Bruce and J. K. Tsotsos, "Saliency , attention , and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, no. 3, pp. 1–24, 2009. doi: 10.1167/9.3.5.

[23] A. Borji, "Saliency prediction in the deep learning era: Successes, limitations, and future challenges," 2019. [Online]. Available: https://arxiv.org/abs/1810.03716

[24] M. Kümmerer, T. S. A. Wallis, and M. Bethge, "DeepGaze II: Reading fixations from deep features trained on object recognition," 2016. [Online]. Available: http://arxiv.org/abs/1610.01563

[25] V. Navalpakkam and L. Itti, "Modeling the influence of task on attention," *Vision Research*, vol. 45, no. 2, pp. 205–231, 2005. doi: 10.1016/j.visres.2004.07.042

[26] A. D. Hwang, E. C. Higgins, and M. Pomplun, "A model of top-down attentional control during visual search in complex scenes," *Journal of Vision*, vol. 9, no. 5, pp. 1–18, 2009. doi: 10.1167/9.5.25.A

[27] A. Borji, D. N. Sihite, and L. Itti, "What/where to look next? Modeling top-down visual attention in complex interactive environments," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 5, pp. 523–538, 2014. [Online]. Available: http://ilab.usc.edu/

[28] A. Kimura, D. Pang, T. Takeuchi, K. Miyazato, J. Yamato, and K. Kashino, "A stochastic model of human visual attention with a dynamic Bayesian network," 2008. [Online]. Available: http://arxiv.org/abs/1004.0085

[29] S. Chikkerur, T. Serre, C. Tan, and T. Poggio, "What and where: A Bayesian inference theory of attention," *Vision Research*, vol. 50, no. 22, pp. 2233–2247, 2010. doi: 10.1016/j.visres.2010.05.013

[30] V. Maljkovic and K. Nakayama, "Priming of pop-out: I. Role of features," *Memory and Cognition*, vol. 22, no. 6, pp. 657–672, 1994. doi: 10.3758/BF03209251

[31] G. Kristjánsson, Árni, Campana, "Where perception meets memory: A review of repetition priming in visual search tasks," *Attention, Perception, & Psychophysics*, vol. 72, pp. 5–18, 2010. doi: 10.3758/APP.72.1.5

[32] J. Theeuwes and E. van der Burg, "On the limits of top-down control of visual selection," *Attention, Perception, and Psychophysics*, vol. 73, no. 7, pp. 2092–2103, 2011. doi: 10.3758/s13414-011-0176-9

[33] Y. C. Tseng, J. I. Glaser, E. Caddigan, and A. Lleras, "Modeling the effect of selection history on pop-out visual search," *PLoS ONE*, vol. 9, no. 3, 2014. doi: 10.1371/journal.pone.0089996

[34] R. Ratcliff, "A theory of memory retreival," *Psychological Review*, vol. 85, no. 2, pp. 59–108, 1978. doi: 10.1037/h0021465

[35] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Company, 1982. ISBN 0-7167-1284-9

[36] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980. doi: 10.1016/0010-0285(80)90005-5

[37] J. Theeuwes and M. Failing, *Attentional Selection: Top-Down, Bottom-Up and History-Based Biases (Elements in Perception)*. Cambridge: Cambridge University Press., 2020. ISBN 9781108891288

[38] D. Matzke and E. J. Wagenmakers, "Psychological interpretation of the ex-gaussian and shifted wald parameters: A diffusion model analysis," *Psychonomic Bulletin and Review*, vol. 16, no. 5, pp. 798–817, 2009. doi: 10.3758/PBR.16.5.798

[39] R. D. Luce, *Response times: Their role in inferring elementary mental organization*. Oxford University Press, 1986.

[40] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985. doi: 10.1007/978-94-009-3833-5

[41] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006. ISBN 9780387310732

[42] D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge: Cambridge University Press, 2012. ISBN 9780511804779

[43] D. M. Endres, E. Chiovetto, and M. A. Giese, "Model selection for the extraction of movement primitives," *Frontiers in Computational Neuroscience*, vol. 7, p. 185, 2013. doi: 10.3389/fncom.2013.00185

[44] H. R. Liesefeld, A. M. Liesefeld, S. Pollmann, and H. J. Müller, "Biasing allocations of attention via selective weighting of saliency signals: behavioral and neuroimaging evidence for the Dimension-Weighting Account," in *In T. Hodgson (Ed.), Processes of Visuospatial Attention and Working Memory*. Springer, 2019. ISBN 9783642209246

[45] J. Theeuwes, L. Bogaerts, and D. van Moorselaar, "What to expect where and when: how statistical learning drives visual selection," *Trends in Cognitive Sciences*, vol. 26, no. 10, pp. 860–872, 2022. doi: 10.1016/j.tics.2022.06.001. [Online]. Available: https://doi.org/10.1016/j.tics.2022.06.001

[46] J. T. Mordkoff and S. Yantis, "An Interactive Race Model of Divided Attention," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 17, no. 2, pp. 520–538, 1991. doi: 10.1037/0096-1523.17.2.520

[47] R. Harald Baayen and P. Milin, "Analyzing reaction times," *International Journal of Psychological Research*, vol. 3, no. 2, p. 12, 2017. doi: 10.21500/20112084.807

[48] M. Santhanagopalan, M. Chetty, C. Foale, S. Aryal, and B. Klein, "Modeling neurocognitive reaction time with gamma distribution," *Proceedings of the Australasian Computer Science Week Multiconference*, 2018. doi: 10.1145/3167918.3167941

[49] J. V. Stone, "Using reaction times and binary responses to estimate psychophysical performance: An information theoretic analysis," *Frontiers in Neuroscience*, vol. 8, 2014. doi: 10.3389/fnins.2014.00035. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnins.2014.00035/abstract

[50] M. D. P. Martin and Fermin, "A theory of reaction time distributions," Tech. Rep., 2008.

[51] R. H. S. Carpenter, "Oculomotor procrastination." *Eye Movements: Cognition and Visual Perception, eds D. F. Fisher, R. A. Monty and J. W. Senders (Hillsdale, NJ: Lawrence Erlbaum)*, pp. 237–246, 1981. doi: 10.4324/9781315437415-19

[52] I. Noorani and R. H. Carpenter, "The LATER model of reaction time and decision," *Neuroscience and Biobehavioral Reviews*, vol. 64, pp. 229–251, 2016. doi: 10.1016/j.neubiorev.2016.02.018

[53] R. Whelan, "Effective analysis of reaction time data," *Psychological Record*, vol. 58, no. 3, pp. 475–482, 2008. doi: 10.1007/BF03395630

[54] L. Christie and R. D. Luce, "Decision structure and time relations in simple choice behavior," *The bulletin of mathematical biophysics*, vol. 18, pp. 89–112, 1956.

[55] W. J. McGill, "Stochastic latency mechanisms," *In D. Luce (ed.), Handbook of Mathematical Psychology, John Wiley & Sons.*, pp. 1–309, 1963.

[56] R. H. Hohle, "Inferred components of reaction times as functions of foreperiod duration," *Journal of Experimental Psychology*, vol. 69, no. 4, pp. 382–386, 1965. doi: 10.1037/h0021740

[57] E. M. Palmer, T. S. Horowitz, A. Torralba, and J. M. Wolfe, "What are the shapes of response time distributions in visual search?" *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 1, pp. 58–71, 2011. doi: 10.1037/a0020747

[58] S. L. Hwang-Gu, Y. C. Chen, S. H. Y. Liang, H. C. Ni, H. Y. Lin, C. F. Lin, and S. S. F. Gau, "Exploring the variability in reaction times of preschoolers at risk of attention-deficit/hyperactivity disorder: An ex-Gaussian analysis," *Journal of Abnormal Child Psychology*, vol. 47, no. 8, pp. 1315–1326, 2019. doi: 10.1007/s10802-018-00508-z

[59] S. L. Hwang Gu, S. S. F. Gau, S. W. Tzang, and W. Y. Hsu, "The ex-Gaussian distribution of reaction times in adolescents with attention-deficit/hyperactivity disorder," *Research in Developmental Disabilities*, vol. 34, no. 11, pp. 3709–3719, 2013. doi: 10.1016/j.ridd.2013.07.025

[60] D. C. Osmon, D. Kazakov, O. Santos, and M. T. Kassel, "Non-Gaussian Distributional Analyses of Reaction Times (RT): Improvements that Increase Efficacy of RT Tasks for Describing Cognitive Processes," *Neuropsychology Review*, vol. 28, no. 3, pp. 359–376, 2018. doi: 10.1007/s11065-018-9382-8

[61] N. Dias, *Eye-Tracking Measures of Attentional Bias in Cocaine*. UT GSBS Dissertations and Theses (Open Access).Paper 446., 2014.

APPENDIX D

THIRD MANUSCRIPT

Sensorimotor processes are not a source of much noise: sensorimotor and decision components of reaction times

Neda Meibodi (meibodi@uni-marburg.de)

Anna Schubö (schuboe@uni-marburg.de)

Dominik Endres (dominik.endres@uni-marburg.de)

Department of Psychology, Philipps-University Marburg, Gutenbergstrasse 18,
35032 Marburg, Germany

Abstract

Statistical descriptions of reaction times are central components of quantitative attention models. It is often assumed that total reaction time is comprised of various components, e.g. sensory delays, decision making and motor execution contributions. We use machine learning to decompose observed total reaction times into sensorimotor and decision components, and evaluate which model assumptions maximize approximate Bayesian model evidence (free energy or evidence lower bound). We find that an inverse Gaussian decision time distribution combined with a very narrow Gaussian sensorimotor distribution can best explain human reaction time data. We also model outliers explicitly by a uniform background distribution. We find that the model assigns a small fraction of datapoints to this outlier distribution.

Keywords: Decision component; Expectation maximization; Free energy; Inverse Gaussian; Mixture models; Reaction time; Sensorimotor component; Visual attention.

Introduction

Reaction time has been widely used as a measure of cognitive processes, e.g. in attention research. It is believed that total observed reaction time (RT) is a sum of different time components. As Luce (1986) mentioned at least five processes may contribute to a total reaction time: physical input transduction into neural spikes, spike transmission to the brain, signal processing and motor programming for the target muscle group (we call this part *decision time*), signal transmission to the muscles and eventually muscle contraction. As it is hard to observe all these components separately, we stack them all – except decision time – and call them the *sensorimotor component* of reaction time. This component is commonly called residual latency (Luce, 1986) or non-decision time (Ratcliff & Tuerlinckx, 2002). During the last decades, there have been many proposals and investigations on how these components combine to yield the final RT distributions.

In some older research (Christie & Luce, 1956; Hohle, 1965), it is reported that RT is a sum of a Gaussian and a exponentially distributed component, where one represents the decision time and another represents the motor component. Consequently, the Ex-Gaussian distribution, which results from convolving these two distributions has been used for modelling RT distributions and cognitive processes (Ratcliff, 1978; Hohle, 1965; Fitousi, 2020; Meibodi, Abbasi, Schubö, & Endres, 2021b) and also psychological disorders (Hwang-Gu et al., 2019). Meibodi et al. (2021b) proposed a model of visual attention which predicts parameters of RT distributions. In that study, an analysis of RTs showed that the ex-Gaussian is a better descriptor than other commonly used distributions, followed by an inverse Gaussian (Meibodi, Abbasi, Schubö, & Endres, 2021a). The authors of that study modelled total RTs without considering a decomposition into separate components, we would like to remedy this shortcoming here. However, the ex-Gaussian has several features that are theoretically not convincing (Schwarz, 2001): first, the Gaussian component has been linked to either the decision or the motor process. Since both processes must take a positive amount of time, a (wide) Gaussian is not a plausible distribution. Second, there is no compelling connection between the distribution parameters and theoretical accounts of the origin of reaction times. Third, the hazard function of the ex-Gaussian is increasing although the best descriptive RT distributions have been reported to have peaked hazard functions (Maddox, Ashby, & Gottlob, 1998).

To address these issues, Schwarz (2001) proposed the ex-Wald distribution for RTs, which is a convolution of an inverse Gaussian with an exponential. Here, the inverse Gaussian describes the decision time, whereas the non-decision component is distributed exponentially (Palmer, Horowitz, Torralba, & Wolfe, 2011). One appealing feature of the ex-Wald is the inverse Gaussian component which models the first passage time distribution of a random walk (Folks & Chhikara, 1978). Such random walks describe quasi-Bayesian sensory evidence accumulation, or drift-diffusion processes. On the other hand, the claim that non-decision time has an exponential distribution seems unjustified. Although the exponential component is commonly interpreted as the effect of a residual process, there are some controversial opinions mentioning that the exponential effect on RT distribution just reflects the search process in visual search tasks (Horowitz & Wolfe, 2003; Palmer et al., 2011).

In (Ratcliff & Tuerlinckx, 2002)'s drift diffusion model (DDM), the parameter T_{er} denotes the time that is spent on processes other than the decision making – such as stimulus encoding, response output and memory access. The parameter has variability to correct the model fits on different data sets under variety of conditions. In this model non-decision time is uniformly distributed (Ratcliff & Tuerlinckx, 2002; Hawkins, Forstmann, Wagenmakers, Ratcliff, & Brown, 2015) although it is mentioned that the true dis-

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

tribution might be normal (Wiecki, Sofer, & Frank, 2013) or skewed. Ratcliff claimed that the shape of the reaction time distribution is primarily determined by the shape of the decision component and the precise shape of the non-decision distribution has a small effect on that as the former has a very large standard deviation (Ratcliff & Tuerlinckx, 2002; Ratcliff & Smith, 2004; Ratcliff, 2006; Ratcliff & McKoon, 2008; Ratcliff & Childers, 2015). One drawback of that model is that the non-decision component happens before and after decision part (Ratcliff & McKoon, 2008) although the result of some studies indicated that these component are intertwined (Evans & Wagenmakers, 2020).

The mean of the non-decision time reported in DDM is about 300 ms with a standard deviation in range 3 to 10 (Ratcliff & Tuerlinckx, 2002). The reported range of non-decision time might differ in other studies based on the experiment, apparatus or participants' attributes. Using simple reaction time (SRT) experiments, previous research has tried to determine how response delay is influenced by features such as: colour of stimuli (Amini Vishteh, Mirzajani, Jafarzadehpour, & Darvishpour, 2019), participants' age (Jain, Bansal, Kumar, & Singh, 2015; Woods, Wyma, Yund, Herron, & Reed, 2015), gender (Dykiert, Der, Starr, & Deary, 2012; Jain et al., 2015), physical activities (Jain et al., 2015) or computer hardware and software (Dodonova & Dodonov, 2013). In a typical SRT study, participants have to press a key as soon as they see the stimulus on the screen (Ulrich & Stapf, 1984). We therefore assume that a SRT contains only a very short decision component and that it is dominated by the sensory and motor processing times. Hence, a SRT approximates the part of a RT which we call the SM component. See Table 1 for an overview of the reported results.

The importance of good RT distribution models is their applicability to statistical analysis and to the modelling of cognitive psychological processes. Most psychologists are interested in decision component of RT and look at the rest of it (commonly called residual latency) as a nuisance variable that should be subtracted from RT (Luce, 1986). However, in addition to decision component analysis, looking at residual latency is also informative. For instance the result of Pedersen, Frank, and Biele (2017) showed that longer RTs in medicated ADHD participants arose because of a strong increase in their non-decision (residual latency) time. Ratcliff, Thapar, and McKoon (2001) found that in some tasks, slower responses of older participants can also be the effect of longer non-decision time.

In this paper we try to disentangle the components of RTs and to recognize outlier responses using a machine learning approach derived from free energy minimization (Friston, Kilner, & Harrison, 2006). We do this with the aim of making attention models, e.g. the one presented by Meibodi et al. (2021b) more interpretable in terms of the underlying psychological processes. We investigate several proposals for the distribution of the SM component: Gaussian, gamma and Laplace. Our motivation for testing the Gaussian distribution is its popularity in previous research, e.g. Christie and Luce (1956); Hohle (1965); Ratcliff and Tuerlinckx (2002) as discussed above. The Laplace distribution has heavier tails rather than the Gaussian and might therefore be less sensitive to extreme SM variations. Both distributions are supported on $\mathbb{R}$ and assign non-zero probability to negative SM components, which is implausible. We therefore experimented with the gamma distribution that has a positive support. Furthermore, since motor output is driven by neuronal spiking activity, its timing would be determined by spike arrival at the neuromuscular synapses. The gamma distribution has been used before to model inter-spike intervals (Ostojic, 2011). In the next section, we will describe the models, followed by a short description of the database used for learning. We then present model comparison results, which indicate that an inverse Gaussian decision time distribution combined with a very narrow Gaussian sensorimotor distribution can best explain human reaction time data. We also model outliers explicitly by a uniform background distribution. We find that the model assigns only a small fraction of datapoints to this outlier distribution. Finally, we discuss the implications of our findings.

Methods

We model an RTs as mixtures of two models ($M = 0$ and $M = 1$) as shown in Figure 1. If $M = 1$ (the response model), then an RT has two components, namely 'decision' and 'sensorimotor (SM)'. If $M = 0$, then the RT is assumed to be an outlier which is drawn from a uniform distribution in range $(0, t_{max})$, i.e. an outlier response has no relationship to the task other than its occurrence before the trial's end at t_{max}. In this case, all we know about the response is that it may happen at any time point in $[t_{min}, t_{max}]$, which is captured by the uniform distribution. For $M = 1$, we assume that the decision component can be viewed as the first passage time in a Wiener diffusion process, which is a model of Bayesian evidence accumulation. Thus, the distribution of the decision component is an inverse Gaussian (Folks & Chhikara, 1978; Schwarz, 2001). The SM component's distribution precise shape has no clear theoretical motivation, hence we try to determine it

Table 1: Mean and standard deviation (SD) of some SRT experiments. The smallest and largest reported mean can be seen in the table for each study. The reported means vary based on between-group differences such as participants' age/gender or stimuli features.

Study	Mean $\pm SD$ (ms)
Amini Vishteh et al. (2019)	207.88 $\pm$7.14
	224.39 $\pm$15.62
Jain et al. (2015)	217.13 $\pm$12.60
	256.36 $\pm$20.34
Woods et al. (2015)	217.9 $\pm$19.5
	239.1 $\pm$28.1

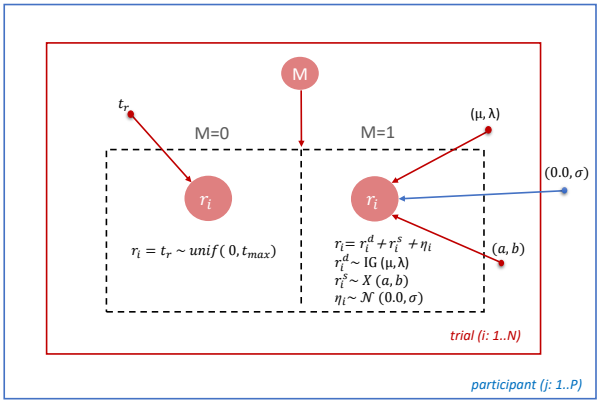

Figure 1: Mixture model of inverse Gaussian and uniform distributions. The blue and red boxes represent participants and trials respectively. P is the number of participants and N is the number of trials for each participant. Each trial's reaction time (r_i) can follow one of the models (M=0 or M=1). When M=0, r_i is a sample of uniform distribution $(0, t_{max})$ and t_{max} is maximum time window that participants had for each trial. In that case, the assumption is that RT can not be decomposed into SM and decision components and it is considered as an outlier. When M=1 then each r_i has 2 main components –decision component (r_i^d) and SM component (r_i^s)– and also noise (η_i) which is normally distributed. The assumption is that (r_i^s) can be a sample from a distribution (X) with parameters a and b. Here, X is either Gaussian, gamma or Laplace which we determine by model comparison.

by model comparison between a Gaussian, a Laplacian and a Gamma distribution. Note that the Gaussian can only be a suitable candidate if it is so narrow that the probability for a negative RT is virtually zero.

We assume that there is one SM distribution per participant, as shown in Figure 1, Thus, each reaction time (r_i) is a sum of a decision component (r_i^d) , SM component (r_i^s) and measurement noise (η_i)

$$\begin{aligned} r_i &= r_i^s + r_i^d + \eta_i \\ \eta_i &\sim \mathcal{N}(0.0, \sigma) \\ r_i^s &\sim X(a,b) \\ r_i^d &\sim IG(\mu, \lambda) \end{aligned} \tag{1}$$

where IG is inv-Gaussian (inverse Gaussian) distribution and X is the SM distribution. We tested the model with gamma, Gaussian and Laplace as the X distribution (see the models comparison in result section). Normally distributed noise (η_i) describes the random effects on the measurement process.

Since exact inference is intractable in this model, we are instead maximizing a lower bound on the expected log likelihood, or evidence lower bound (ELBO) (Bishop, 2006) a.k.a. free energy (Friston, 2003). The ELBO of our model is

$$\begin{aligned} L = \int dr_i^s \int dr_i^d \sum_{M_i=0}^{1} \sum_{i=1}^{N} \Big[M_i \big[\log p(r_i|r_i^s, r_i^d, \sigma) + \\ - \log\big(q(r_i^d|\ddot{\theta}_i^d)/p(r_i^d|\theta^d)\big) - \log\big(q(r_i^s|\ddot{\theta}_i^s)/p(r_i^s|\theta^s)\big)\big] \\ + (1-M_i)\log p(r_i) - \log\big(q(M_i)/p(M_i)\big) \\ \Big] q(M_i)\ \big(q(r_i^s|\ddot{\theta}_i^s)\ q(r_i^d|\ddot{\theta}_i^d)\big)^{M_i} \end{aligned} \tag{2}$$

where N is the number of trials for each participant, M is the model type, q is the variational posterior distribution and θ includes inv-Gaussian parameters for each r_i when $M_i = 1$. θ^s and θ^d are prior parameters on the decision component (μ, λ) and SM component (a, b) and $\ddot{\theta}_i^s$, $\ddot{\theta}_i^d$ are posterior parameters on the same components for each trial (i). When $M = 0$, the RT can not be decomposed, which is modelled by the uniform distribution $p(r_i)$ (see Figure 1).

Using the usual definition of the Kullback-Leibler divergence (KL) between distributions q and p

$$KL(q(x)|p(x)) = \int dx q(x)\big[\log(q(x)/p(x))\big] \tag{3}$$

we can rewrite Eq 2 as

$$\begin{aligned} L = \sum_{i=1}^{N} \Big[q(M_i=1)\big[\langle \log p(r_i|r_i^s, r_i^d, \sigma)\rangle_{q(r_i^s|\ddot{\theta}_i^s)\ q(r_i^d|\ddot{\theta}_i^d)} \\ - KL(q(r_i^s|\ddot{\theta}_i^s)|p(r_i^s|\theta^s)) - KL(q(r_i^d|\ddot{\theta}_i^d)|p(r_i^d|\theta^d))\big] \\ + q(M_i=0)\ \log p(r_i) - KL(q(M_i)|p(M_i))\Big] \end{aligned} \tag{4}$$

where $\langle p(r_i|r_i^s, r_i^d, \sigma)\rangle_{q(r_i^s|\ddot{\theta}_i^s)\ q(r_i^d|\ddot{\theta}_i^d)}$ is the expectation of the conditional probability with respect to $q(r_i^s|\ddot{\theta}_i^s)$ and $q(r_i^d|\ddot{\theta}_i^d)$. See Appendix for the derivations of this term. We assume that the variational posteriors are from the same family of distributions as the respective prior.

We then optimize the bound with respect to q parameters: $\ddot{\theta}_i^d$ and $\ddot{\theta}_i^s$. These optimizations, which we carry out in an alternating fashion, can be viewed as the E and M steps of a variational expectation maximization (EM) algorithm (Barber, 2012). In each E-step, for fixed parameters $(\ddot{\theta}_i^s, \ddot{\theta}_i^d)$ we find the distribution $q(M)$ which maximizes Eq 4 and in each M-step, we find $\ddot{\theta}^d$ and $\ddot{\theta}^s$ that maximize Eq 4 while $q(M)$ is fixed. Additionally, we also update the prior parameters at the end of each M-step. We did not choose to equip the prior parameters with a hyperprior, because we expect them to be well determined by the data ($N > 1000$ trials per participant). Figure 2 shows a flow chart of the optimization steps. We implemented the model with Pytorch (1.10.1) in Python (3.9.7) using the Adam optimizer. For more information about learning rates and iteration steps, see the code at: `http://dx.doi.org/10.17192/fdr/88`.

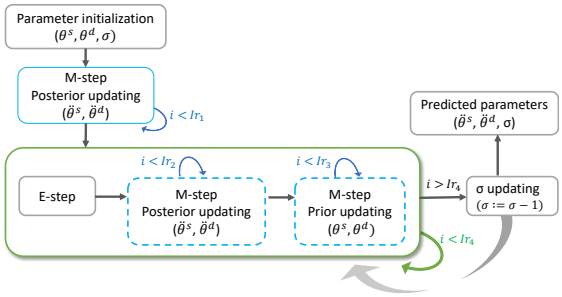

Figure 2: Flowchart of the variational E-M algorithm used for model optimization. θ^s, θ^d and σ are the parameters of SM component, decision component and the standard deviation of the noise, respectively (see Eq 4). Each blue box is an optimization step and blue arrows indicate that the parts are iterated to converge – with negative ELBO as the loss functions. Ir is iteration rate and its value differs among optimization parts ($Ir_1 \cdots Ir_4$). The whole green box is also iterated to converge. Additionally, after each σ updating (to seek the smallest possible standard deviation for the noise function) the whole green box is iterated again. $\ddot{\theta}^s$, $\ddot{\theta}^d$ and σ are predicted posterior parameters for SM component, decision component and standard deviation of the noise function, respectively. In the first M-step (the one out of the green box) we assume that all data points belong to $M = 1$ (the response model), to obtain initial parameter estimates, since we expect a only a small fraction of outliers.

Database

We optimized the model on a RT database of a visual attention experiment from (Feldmann-Wüstefeld, Uengoer, & Schubö, 2015) (the first experiment out of four) which includes two different types of intermixed tasks. Participants were presented with eight elements on an imaginary circle around the fixation point. In one task, they had to responded to either the shape or the colour singleton based on their group membership. In the other task , both groups of participants responded to the orientation of a line embedded in the shape singleton while they had to ignore a color distractor in half of the trials. The target of the experiment was investigating the role of selection history (Awh, Belopolsky, & Theeuwes, 2012) on selective visual attention. The participants (11 males and 17 females) were in 18-32 age range and all but two were right handed. Each participant responded to 1024 trials. For more information about the experiment, see the main reference (Feldmann-Wüstefeld et al., 2015). Meibodi et al. (2021b) proposed a model for these data which assumes an ex-Gaussian RT distribution, we aim to replace this model assumption by a more theoretically motivated one.

Results and discussion

As mentioned in previous section, we tested three versions of the model with different X distributions (Gaussian, gamma or Laplace) (see Eq 1 and Figure 1). We selected the prior on these distributions' parameters in a way that matches the reported means and standard deviations in other SRT studies(see Table 1). We chose the prior on inv-Gaussian parameters based on a preliminary analysis of the data which assumed that the SM component is a constant. More specifically, priors parameter values are

$$\begin{aligned} r^d &\sim IG(\mu = 500.0, \lambda = 10000.0) \\ \eta &\sim \mathcal{N}(0.0, \sigma = 12.0) \\ r^s &\sim Laplace(m = 200.0,\ std = 9.8)\ or \\ &\quad Gaussian(m = 200.0,\ std = 10.0)\ or \\ &\quad Gamma(m = 200.0,\ std = 10.0). \end{aligned} \tag{5}$$

The model was then fitted to the data of each of the 28 participants. The number of participants is within the range typically used in mixed model repeated measurement designs (Feldmann-Wüstefeld et al., 2015). For each participant the final free energy is computed (the results are plotted in Figure 3a for all models) and the sum over all participants is used for model comparison (see Table 2). Smaller free energy values indicate better fits (higher ELBO). Thus, the best model is the one with Gaussian SM distribution. The results of the Laplace model are very close to the Gaussian. The mean of outliers over all participants is also shown in Table 2. The models label a very similar proportion of trials as outliers, independent of the choice of SM distribution, as can be seen in Figure 3b. This closeness might be due to having similar decision components in all versions of the model. The decision component (inv-Gaussian) has a much bigger variance than the narrow SM components– and it is therefore driving the outlier determination. As explained in the methods section, outliers are RTs which can not be separated to decision and SM components by the model. So they are sampled from a uniform distribution (M=0). Thus, an outlier is either a very fast or a very slow response. For two example participants, Figure 4 illustrates that which part of the data is considered as an outlier by our model. The criterion in these plots is

Table 2: Model comparison results for different SM distributions. 'FE' is free energy (sum over all participants for each model), smaller values indicate a better RT database fit. 'Outliers' shows the mean fraction of outliers over all participants. For each participant the amount of outliers is the sum over the outlier posterior distribution ($q(M = 0)$). 'θ^s' includes the updated prior parameters (mean and standard deviation) of the SM distribution which are optimized by the model. The reported values are the grand means of the means and standard deviations over all participants.

SM distribution	FE	Outliers	θ^s (mean, std)
Gaussian	187387.66	1.52%	199.58, 0.37
Laplace	187685.74	1.53%	199.59, 0.40
Gamma	191684.43	1.68%	199.50, 0.52

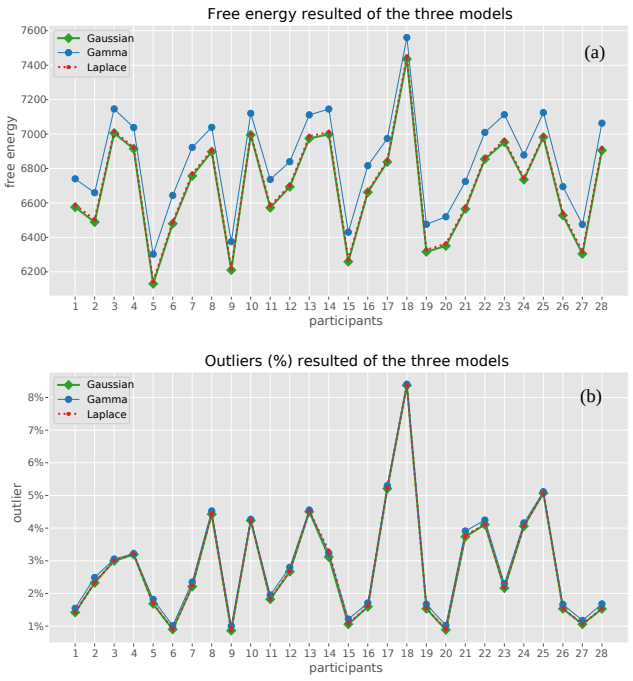

Figure 3: Results per participants. For 28 participants see the free energy values (a) and the percentage of outliers (b) resulting from the optimization of three different SM distribution models.

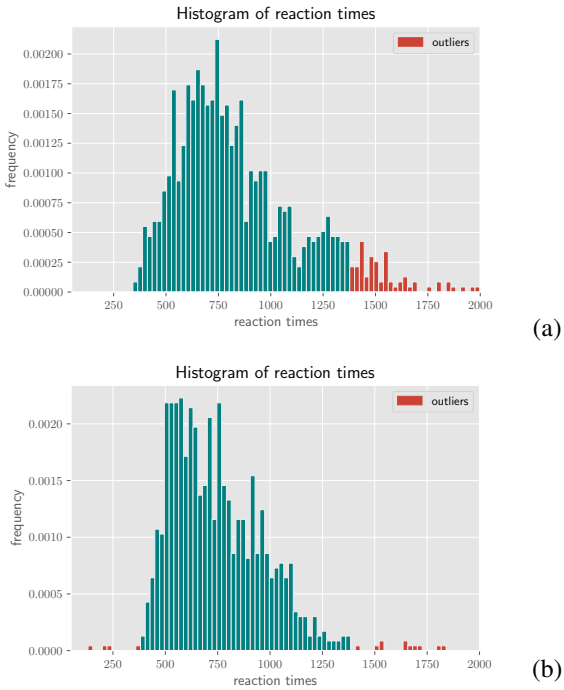

Figure 4: Outliers. Histograms of RTs with predicted outliers (marked in red) by the model for the 18th and the 3rd participant. These participants have relatively: the highest numbers of outliers (a) and an average numbers of outliers –with very short and long RTs (b).

$q_i(M=0) > 0.6$ which means r_i is labelled as an outlier if it belongs to $M=0$ with a probability of more than 0.6.

For each trial, the model predicts posterior parameters on both decision and SM components through minimizing free energy and afterwards priors are updated (these steps are shown in Figure 2). The final prior parameters of SM component (θ^s) for each version of the model can also be seen in Table 2. These predicted parameters are close in mean and all are very narrow distributions (see Figure 5).

Subtracting both the expected SM component (r_i^s) and the noise component (η_i) from the RTs ($r_i^d = r_i - r_i^s - \eta_i$), we obtain the expected decision component. The inverse Gaussian distribution fits better on this component rather than on the total RT data, indicating that the Wiener process assumption might be justified. However, this assumption should be investigated more in future studies. For an illustration, see Figure 6: here, the inverse-Gaussian fits to the total RT data (orange histograms) are worse than the fits to the expected decision components only (in green) for two participants (6th and 18th). These participants have the smallest and the largest numbers of outliers (see and compare their outliers in Figure 3 b). The plots (Figure 6) show that the model works well in either case. In addition, the best distribution for the measurement noise is $\mathcal{N}(0.0, 2.0)$ which is obtained by updating σ at the end of the optimization iterations as shown in Figure 2. Finally, for each participant is possible to reverse the process and reconstruct the RTs from the posteriors. In this case, the mean of reconstruction error for each participant is less than 8.5 ms.

Conclusion

The role and importance of different RT components in shaping the total RT distribution has long been a matter of question in cognitive modelling. Quantitative models, such as the ones proposed in this paper, can be helpful in comparing the predictions of different theoretical accounts of RTs objectively and disentangling the components. Moreover, different lines of research might be interested in different components such as the effect of brain disorders on decision making (Herz, Bogacz, & Brown, 2016) versus motor responses (Low, Miller, & Vierck, 2002).

The purpose of the modelling reported in this paper was to investigate if machine learning methods can help to disentangle a RT distribution into two main components of decision time and sensorimotor time. The motivation for our research was a previous study by Meibodi et al. (2021b) which presented an algorithmic model of selection history effects without a solid theoretical foundation for the chosen RT distribution. We are now in a position to remedy this issue. We expect that our proposed model will be useful whenever RT components need to be extracted in cognitive RT modeling.

The results showed that the final predicted SM distributions are very narrow which is comparable with the assumption in Ratcliff diffusion model: non-decision component might be

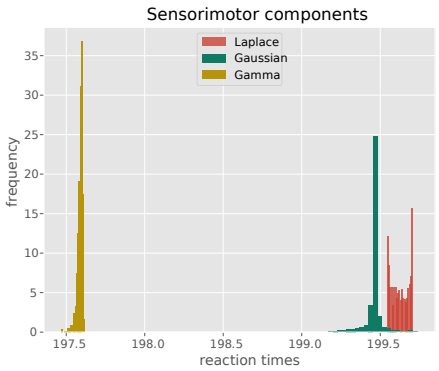

Figure 5: Posterior sensorimotor (SM) components in three versions of the model for a random participant. In these models, it was assumed that SM distributions might be Gaussian, gamma or Laplace. The results show very narrow distributions under all assumptions. Gaussian fits best, see also table 2.

sampled from any distribution and the shape of it can not influence the final RT distribution as the decision part has a very large standard deviation and the other one has a small one (Ratcliff & Childers, 2015). The predicted mean of this distribution is in range of 199.58 ± 0.37 by the best fitted model. The best fitted model is the version which assumes the SM component is Gaussian. The predicted mean has a close range to some simple reaction time experiments results (see Table 1).

The model can successfully label an acceptable number of extreme-valued RTs as outliers. Importantly, in our approach this labelling is driven by the model assumptions and the labels will therefore be internally consistent with the model's predictions, unlike more traditional methods for outlier labels based e.g. on standard deviation measurements. This property might be useful for the principled detection of inattentive participants, e.g. in ADHD or autism studies, where a larger proportion of outliers is to be expected.

Appendix

$$\begin{aligned}
&\langle \log p(r_i|r_i^s, r_i^d, \sigma)\rangle_{q(r_i^s|\ddot{\theta}_i^s)\ q(r_i^d|\ddot{\theta}_i^d)} = \\
&-\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\left\langle (r_i - (r_i^d + r_i^s))^2 \right\rangle_q = \\
&-\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\Big[r_i^2 - 2r_i\langle r_i^d\rangle - 2r_i\langle r_i^s\rangle \\
&+2\langle r_i^s\rangle\langle r_i^d\rangle + \langle r_i^{d2}\rangle + \langle r_i^{s2}\rangle\Big]
\end{aligned} \tag{6}$$

Using the definition of variance $\langle r_i^{d2}\rangle = \langle r_i^d\rangle^2 + var(r_i^d)$, the above equation can be rearranged

$$\begin{aligned}
&\langle \log p(r_i|r_i^s, r_i^d, \sigma)\rangle_{q(r_i^s|\ddot{\theta}_i^s)\ q(r_i^d|\ddot{\theta}_i^d)} = \\
&-\frac{1}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\Big[(r_i - \langle r_i^d\rangle - \langle r_i^s\rangle)^2 \\
&+var(r_i^d) + var(r_i^s)\Big]
\end{aligned} \tag{7}$$

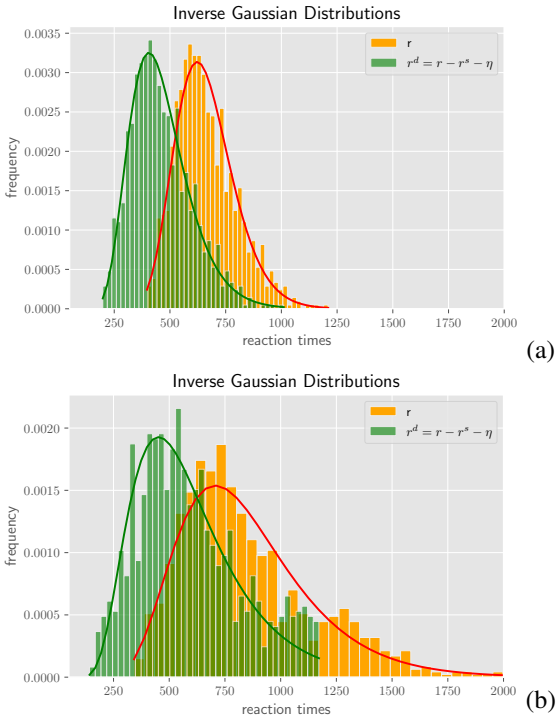

Figure 6: Inverse Gaussian distributions for the 6th participant (a) and the 18th participant (b). The orange histograms contain the total RTs (r), red curves fitted by maximizing log-likelihood. The green histograms show the expected decision components (r^d) after subtracting the SM components (r^s) and noises (η) and also discarding the outliers ($q(M = 0) > 0.6$). Parameters of the green densities are updated priors which are predicted by the model for each participant.

Acknowledgments

This work was supported by the DFG SFB-TRR 135 (222641018) "Cardinal Mechanisms of Perception", projects C6 and B3, and "The Adaptive Mind", funded by the Excellence Program of the Hessian Ministry for Science and the Arts.

References

Amini Vishteh, R., Mirzajani, A., Jafarzadehpour, E., & Darvishpour, S. (2019). Evaluation of simple visual reaction time of different colored light stimuli in visually

normal students. *Clinical Optometry*, *11*, 167–171. doi: 10.2147/OPTO.S236328

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. doi: 10.1016/j.tics.2012.06.010

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge: Cambridge University Press. doi: 10.1017/cbo9780511804779

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Christie, L., & Luce, R. D. (1956). Decision structure and time relations in simple choice behavior. *The bulletin of mathematical biophysics*, *18*, 89–112.

Dodonova, Y. A., & Dodonov, Y. S. (2013). Is there any evidence of historical slowing of reaction time? No, unless we compare apples and oranges. *Intelligence*, *41*(5), 674–687. doi: 10.1016/j.intell.2013.09.001

Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). Sex differences in reaction time mean and intraindividual variability across the life span. *Developmental Psychology*, *48*(5), 1262–1276. doi: 10.1037/a0027550

Evans, N. J., & Wagenmakers, E.-J. (2020). Evidence Accumulation Models: Current Limitations and Future Directions. *The Quantitative Methods for Psychology*, *16*(2), 73–90. doi: 10.20982/tqmp.16.2.p073

Feldmann-Wüstefeld, T., Uengoer, M., & Schubö, A. (2015). You see what you have learned. Evidence for an interrelation of associative learning and visual selective attention. *Psychophysiology*, *52*(11), 1483–1497. doi: 10.1111/psyp.12514

Fitousi, D. (2020). Linking the Ex-Gaussian Parameters to Cognitive Stages: Insights from the Linear Ballistic Accumulator (LBA) Model. *The Quantitative Methods for Psychology*, *16*(2), 91–106. doi: 10.20982/tqmp.16.2.p091

Folks, J. L., & Chhikara, R. S. (1978). The Inverse Gaussian Distribution and Its Statistical Application-A Review. *Journal of the Royal Statistical Society*, *40*(3), 263–289. doi: 10.1111/j.2517-6161.1978.tb01039.x

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325–1352. doi: 10.1016/j.neunet.2003.06.005

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology Paris*, *100*(1-3), 70–87. doi: 10.1016/j.jphysparis.2006.10.001

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E. J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience*, *35*(6), 2476–2484. doi: 10.1523/JNEUROSCI.2410-14.2015

Herz, D. M., Bogacz, R., & Brown, P. (2016). Neuroscience: Impaired Decision-Making in Parkinson's Disease. *Current Biology*, *26*(14), R671–R673. doi: 10.1016/j.cub.2016.05.075

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*(4), 382–386. doi: 10.1037/h0021740

Horowitz, T. S., & Wolfe, J. M. (2003). Memory for rejected distractors in visual search? *Visual Cognition*, *10*(3), 257–298. doi: 10.1080/13506280143000005

Hwang-Gu, S. L., Chen, Y. C., Liang, S. H. Y., Ni, H. C., Lin, H. Y., Lin, C. F., & Gau, S. S. F. (2019). Exploring the variability in reaction times of preschoolers at risk of attention-deficit/hyperactivity disorder: An ex-Gaussian analysis. *Journal of Abnormal Child Psychology*, *47*(8), 1315–1326. doi: 10.1007/s10802-018-00508-z

Jain, A., Bansal, R., Kumar, A., & Singh, K. (2015). A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *International Journal of Applied and Basic Medical Research*, *5*(2), 124. doi: 10.4103/2229-516x.157168

Low, K. A., Miller, J., & Vierck, E. (2002). Response slowing in Parkinson's disease: A psychophysiological analysis of premotor and motor processes. *Brain*, *125*(9), 1980–1994. doi: 10.1093/brain/awf206

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford University Press.

Maddox, W. T., Ashby, F. G., & Gottlob, L. R. (1998). Response time distributions in multidimensional perceptual categorization. *Perception and Psychophysics*, *60*(4), 620–637. doi: 10.3758/BF03206050

Meibodi, N., Abbasi, H., Schubö, A., & Endres, D. (2021a). Distracted by previous reward: integrating selection history, current task demands and saliency in a computational model. *PsyArXiv*. doi: 10.31234/OSF.IO/MBE5A

Meibodi, N., Abbasi, H., Schubö, A., & Endres, D. (2021b). A model of selection history in visual attention. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*. Retrieved from `https://escholarship.org/uc/item/3m33h9h7`

Ostojic, S. (2011). Interspike interval distributions of spiking neurons driven by fluctuating inputs. *Journal of Neurophysiology*, *106*(1), 361–373. doi: 10.1152/jn.00830.2010

Palmer, E. M., Horowitz, T. S., Torralba, A., & Wolfe, J. M. (2011). What are the shapes of response time distributions in visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 58–71. doi: 10.1037/a0020747

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin and Review*, *24*(4), 1234–1251. doi: 10.3758/s13423-016-1199-y

Ratcliff, R. (1978). A theory of memory retreival. *Psychological Review*, *85*(2), 59–108. doi: 10.1037/h0021465

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, *53*(3), 195–237. doi: 10.1016/j.cogpsych.2005.10.002

Ratcliff, R., & Childers, R. (2015). Individual Differences and Fitting Methods for the Two-Choice Diffusion Model

of Decision Making. *Decision*, *2*(4), 237–279.

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. doi: 10.1016/j.biotechadv.2011.08.021.Secreted

Ratcliff, R., & Smith, P. L. (2004). A Comparison of Sequential Sampling Models for Two-Choice Reaction Time. *Psychological Review*, *111*(2), 333–367. doi: 10.1037/0033-295X.111.2.333

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, *16*(2), 323–341. doi: 10.1037/0882-7974.16.2.323

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review*, *9*(3), 438–481. doi: 10.3758/BF03196302

Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments, and Computers*, *33*(4), 457–469. doi: 10.3758/BF03195403

Ulrich, R., & Stapf, K. H. (1984). A double-response paradigm to study stimulus intensity effects upon the motor system in simple reaction time experiments. *Perception & Psychophysics*, *36*(6), 545–558. doi: 10.3758/BF03207515

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical bayesian estimation of the drift-diffusion model in Python. *Frontiers in Neuroinformatics*, *7*. doi: 10.3389/fninf.2013.00014

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience*, *9*. doi: 10.3389/fnhum.2015.00131

APPENDIX E

CURRICULUM VITAE

The curriculum vitae is deleted in order to data privacy.

APPENDIX F

ZUSAMMENFASSUNG IN DEUTSCHER SPRACHE

In den letzten zwei Jahrzehnten wurden zahlreiche Modelle der visuellen Aufmerksamkeit vorgeschlagen. Forscher in verschiedenen Disziplinen, wie der Psychologie und Ingenieurswissenschaften sind daran interessiert, die menschlichen Wahrnehmungsmechanismen zu verstehen und/oder Algorithmen zu entwickeln, welche die Aufmerksamkeitsprozesse für bestimmte Anwendungen (z.B. Robotik) nachahmen.

In dieser Dissertation habe ich den Einfluss von Lernerfahrungen auf die Aufmerksamkeitssteuerung modelliert. Das vorgestellte algorithmische Modell verknüpft visuelle Reize mit den Reaktionszeiten von Versuchspersonen in einem Aufmerksamkeitsexperiment. Diese Dissertation besteht aus drei Studien.

In **der ersten Studie** wurde die Rolle der Selektionshistorie –als der Effekt des Lernens aus der Übungsphase des Experiments auf die Hauptphase– untersucht. Ich habe auch auf Dimensionsebene (z.B. Farbe und Form) und auf Merkmalsebene (z.B. blau und rot) getestet. Die Ergebnisse zeigten, dass die Version des Modells, die neben Stimulus-getriebenen (bottom-up) und zielgetriebenen (top-down) Kontrollmechanismen die Auswahlhistorie (auf Merkmalsebene) beinhaltet, am besten für eine quantitative Beschreibung

der Reaktionszeiten der Versuchspersonen geeignet ist.

In **der zweiten Studie** untersuchte ich die Bedeutung des Intertrial Priming –die Auswirkung eines früheren Trials auf die aktuelle– sowie die Bedeutung der Merkmale (Farbe, Form oder Ausrichtung) in den Modellvorhersagen. Es konnte gezeigt werden, dass durch die Einbeziehung des Effekts des Intertrial Priming eine bessere Beschreibung des Verhaltens erreicht werden kann. Außerdem verschlechtert das Ausschließen eines der Merkmale die Modellvorhersagen.

In **der dritten Studie** schlug ich ein Modell zur Aufteilung von Reaktionszeiten –in Entscheidungs und sensomotorische Komponenten– als Voraussetzung für die RT-Modellierung vor. Diese Studie wird uns helfen, genauere Aufmerksamkeitsmodelle einzuführen. Darüber hinaus kann es kognitive Studien unterstützen, um die Wirkung bestimmter Faktoren (z.B. Alter und psychische Störungen) auf entweder dem motorischen System oder der Entscheidungsfindung besser zu verstehen.

Das vorgeschlagene Aufmerksamkeitsmodell (in der ersten und zweiten Studie) ist eines der ersten Modelle, das den Effekt der Selektionsgeschichte auf die Lenkung der Aufmerksamkeit beinhaltet. Dieses Modell kann die Unterschiede zwischen den Gruppen erfassen, bei denen jede Gruppe von Teilnehmern eine andere Lernerfahrung hatte. Das Modell berücksichtigt die Gesamtreaktionszeiten jedes Teilnehmers. Aber Aufmerksamkeit kann Reaktionszeiten beeinflussen, indem sie verschiedene kognitive Prozesse beeinflusst. Die dritte Studie führt daher eine Methode ein, die uns hilft, jeden Prozess (und seine relevante Reaktionszeitkomponente) unabhängig voneinander zu betrachten.

APPENDIX G

EIGENSTÄNDIGKEITSERKLÄRUNG

Hiermit versichere ich, dass ich die vorliegende Dissertation selbständig, ohne unerlaubte Hilfe Dritter angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Dritte waren an der inhaltlich-materiellen Erstellung der Dissertation nicht beteiligt; insbesondere habe ich hierfür nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden. Mit dem Einsatz von Software zur Erkennung von Plagiaten bin ich einverstanden.

Neda Meiboi .....