

**Measures and models of covert visual attention  
in neurotypical function and ADHD**

by

Andra Mihali

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Center for Neural Science  
New York University  
May 2018

---

Wei Ji Ma

© Andra Mihali

All Rights Reserved, 2018

## Acknowledgements

I cannot thank enough my advisor, Wei Ji Ma, for taking a chance on me, for patiently shaking into precision my noisy thoughts, for the tremendous help both academic and personal, with every aspect of each project. Also, thank you for the durable meta-lessons. I know I will be unpacking for years to come everything you taught me, both explicitly and implicitly, through your example.

I want to thank my committee members, Marisa Carrasco, Eero Simoncelli and Roozbeh Kiani and my external reviewer Susana Martinez-Conde, for their insightful suggestions (some of which I'm still mulling over) and constant inspiration. I am thankful for working with Allison Young, Lenard Adler and Michael Halassa, without whom the ADHD study from Chapter 3 would not have been possible. Ili Ma and two anonymous reviewers provided useful comments on an earlier draft of the manuscript Chapter 3 is based on. In addition, I thank Terry Leon, Michael Silverstein, Saima Milli, Jonathan Yuh for assistance with protocol preparation, participant recruitment and clinical data management.

The majority of the work in Chapter 4 represents joint work with Bas van Opheusden, whose brilliant ideas and enthusiasm got me destuck from local minima countless times. We are grateful to Martina Poletti and Michele Rucci who shared their DPI dataset with us, which allowed us to validate our algorithm. Additionally, this work has benefitted from the comments of Rich Krauzlis and two anonymous reviewers.

Sometimes, it takes a village. The Ma lab (family) has been enormously helpful, both sci-

entifically and personally. One person's problem became everybody's concern and the collective efforts lead to a better solution. I want to especially acknowledge Emin and Luigi, grad students and friends Aspen, Bas and Will (whose positive influence is visible in several figures, scripts, as well as this template) and also Maija (who once helped me troubleshoot an experiment until 1 am), Yanli, Zhiwei, Gianni, Zahy and Aditi. I felt lucky to have had a wonderful and supportive cohort, a few of who became my close friends: Christina, Emily, Kristina, Caro, Robert, Evan, Oli, Alex, Brit, Evelyn, Ain. I am grateful to my first mentor, Stephen Rayport, whose lessons resonated even as I changed fields. Special thanks to our DGS's Mike and Alex for all the concrete, as well as emotional support during a difficult transition. Many thanks to Wendy, Shihpi and Andre for early mentorship. Thanks to Jess, Amala and Erik, and the HPC staff, especially Shenglong Wang. More broadly, I want to thank everybody in the theory suite, and CNS and Psych departments, students and professors, for creating an inspiring environment.

Many thanks Angela and Lyuda, for 10 years of friendship and understanding. Your perspectives on life always brightened my days, all the more so during tough times. Thanks to my dear roommates Milo and Belanna, and former (though forever in spirit) roommate Gillian. Thanks to several other friends, some far away in (mind) space, for treasured past times and/or hopes for the future: Liana, Moni, Lore, Ion, Nic, Nicole, Alex, Veronica, Sumati, Maddy, Pat, Celia, Dan, Vlad, Dragos, Andrei, Cristina, Shai, Gecia, Matt, Scott, Natasha, Delia, Millie, Anton. Of course, I have the deepest gratitude towards my family, especially my parents, Cristina and Vasile Mihali, for unwavering support.

Funding was provided by the NYU MacCracken Fellowship and the following grants: R01EY020958 from the National Eye Institute and W911NF1410476 from the Army Research Office, to Wei Ji Ma, and Applied Research Support Fund (ARSF) 53101 from NYU SoM.

## Preface

Chapter 2 is being prepared for journal submission. Complementary work not presented here, with Aditi Singh as first author, is also in preparation. All publications based on this thesis have or will have Wei Ji Ma as the senior author.

The work described in Chapter 3 has been done in collaboration with Allison G Young as co-first author, as well as Lenard A. Adler and Michael H. Halassa. A manuscript based on this work is under revision at *Computational psychiatry* (preprint (Mihali et al., 2017b)).

The Bayesian microsaccade detection method described in Chapter 4 has been developed with Bas van Opheusden as co-first author and published in *Journal of Vision* (Mihali et al., 2017a). Its application to ADHD entails a preliminary analysis of the fixational eye movements collected while participants performed the experiment described in Chapter 3. A manuscript detailing this work is in the early stages of preparation.

## Abstract

Covert attention allows us to prioritize relevant objects from the visual environment without directing our gaze towards them. Attention affects the quality of perceptual representations, quality which can be quantified with precision (or its inverse, variability) parameters in simple psychophysical models that capture the relationship between stimulus strength and an observer's behavior. Two main types of attention, divided and selective, have been studied in the recent decades with two corresponding classic paradigms, visual search and visual spatial orienting.

In this thesis, we developed variants of these tasks to address questions related to visual attention, in neurotypicals and ADHD. In addition to precision-related parameters derived from behavior, we measured the observers' fixational eye movements, developed a new algorithm to detect microsaccades and explored their possible role as an oculomotor correlate of precision.

In a first investigation, we built upon a paradigm designed to increase the chances of probing divided attention. Specifically, we extended a visual search task with heterogeneous distractors and explored the effects on performance of set size, task - detection and localization, time (perception and memory) and space. An optimal observer model with a variable precision encoding stage and an optimal decision rule was able to capture behavior in a task more naturalistic than target detection, namely target localization. Performance decreased with the set size of the search array for both detection and localization; so did precision. As expected, precision was higher in the perception condition relative to the memory condition. We found the same pattern of results in a second experiment with visual search arrays with reduced stimulus spacing; additionally, observers achieved com-

parable precision parameters, albeit with increased reaction times.

The nature of the attentional impairment in ADHD has been elusive. By using a new task that combines visuo-spatial orienting with feature dimension switch between orientation and color, we found an increased perceptual variability parameter in the ADHD group, which was correlated with an executive control metric. A classifier based on perceptual variability yielded high diagnosis accuracy. These results suggest that using basic psychophysical paradigms to capture encoding precision of low-level features deserves further study in ADHD, especially in conjunction with attention and executive function.

Measures of covert attention have included aspects of fixational eye movements, especially microsaccades. Inferences about the roles of microsaccades in perception and cognition depend on accurate detection algorithms. By using a new hidden semi-Markov model to capture sequences of microsaccades amongst drift and an inference algorithm based on this model, we found that microsaccades were more robustly detected under high measurement noise from the eye tracker. Applying this algorithm to the eye movement traces of ADHD and Control participants, we found a correlation between post-stimulus microsaccade rate and the perceptual variability parameter, suggesting a potential oculomotor mechanism for the less precise perceptual encoding in ADHD.

We conclude that by using and developing variants of visual attention paradigms, psychophysical models and oculomotor measurements, we can enhance our understanding about the brain processes in health and disease.

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>Preface</b>	<b>v</b>
<b>Abstract</b>	<b>vi</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Visual attention . . . . .	2
1.2 Visual search and the feature integration theory of attention . . . . .	3
1.3 Visual search, signal detection theory and divided attention . . . . .	4
1.4 Selective attention . . . . .	6
1.5 Attention-deficit/hyperactivity disorder (ADHD) . . . . .	8
1.6 Microsaccades . . . . .	9
1.7 Dissertation outline . . . . .	12
<b>2 Visual search with heterogenous distractors</b>	<b>14</b>

2.1	Introduction . . . . .	15
2.1.1	Structure of this chapter . . . . .	19
2.2	Experimental methods . . . . .	20
2.2.1	Tasks and experiments overview . . . . .	20
2.2.2	Experiments . . . . .	21
2.3	Models . . . . .	25
2.3.1	Optimal-observer models . . . . .	25
2.3.2	Modeling methods . . . . .	32
2.4	Results Experiment 1 . . . . .	34
2.4.1	Target localization . . . . .	34
2.4.2	Target detection . . . . .	42
2.4.3	Target localization and target detection . . . . .	44
2.5	Results of Experiment 2: effect of stimulus spacing . . . . .	47
2.5.1	Target localization . . . . .	47
2.5.2	Target detection . . . . .	52
2.5.3	Target localization and target detection . . . . .	54
2.6	Discussion . . . . .	56
2.7	Supplementary . . . . .	60
2.7.1	Calculation of quantile bins for the psychometric curves . . . . .	60

### **3 Perception, selective attention and task-switching in neurotypicals and ADHD 62**

3.1	Introduction . . . . .	62
3.2	Methods . . . . .	65
3.2.1	Approach . . . . .	65
3.2.2	Experimental methods . . . . .	69
3.2.3	Statistical analyses . . . . .	73
3.2.4	Parameter fitting . . . . .	75
3.3	Results . . . . .	77
3.3.1	Task-irrelevant motor output (TIMO) . . . . .	77
3.3.2	Reaction times . . . . .	79
3.3.3	Psychometric curve parameters . . . . .	80
3.3.4	Correlations across metrics . . . . .	84
3.3.5	Classification of participants . . . . .	85
3.4	Discussion . . . . .	86
3.4.1	Possible lower-level neural correlates of behavioral variability in ADHD . . . . .	90
3.4.2	Perceptual variability as a candidate diagnosis marker for ADHD . . . . .	91
3.5	Supplementary . . . . .	92
3.5.1	Demographic and clinical information . . . . .	92
3.5.2	Further characterization of responses . . . . .	94
3.5.3	Further information on reaction times . . . . .	96
3.5.4	Further information on psychometric curves and parameters . . . . .	98
3.5.5	Effect of eye tracking . . . . .	104

3.5.6	Breakdown of correlations . . . . .	104
3.5.7	Prediction of clinical variables . . . . .	110
<b>4</b>	<b>Bayesian microsaccade detection with an application to ADHD</b>	<b>112</b>
4.1	Introduction . . . . .	112
4.1.1	Microsaccades in ADHD . . . . .	113
4.1.2	Microsaccade detection methods . . . . .	115
4.1.3	Bayesian methods for saccade and microsaccade detection . . . . .	116
4.1.4	Our approach . . . . .	117
4.1.5	Structure of this chapter . . . . .	118
4.2	Methods . . . . .	118
4.2.1	Generative model . . . . .	120
4.2.2	Inference of the eye state time series . . . . .	124
4.2.3	Approximate inference . . . . .	125
4.2.4	Alternative algorithms . . . . .	131
4.2.5	ADHD microsaccade time course analysis . . . . .	133
4.2.6	Eyelink experimental methods . . . . .	133
4.2.7	DPI Experimental methods . . . . .	134
4.3	Results . . . . .	135
4.3.1	Comparison of BMD and other algorithms on simulated data according to our generative model . . . . .	135
4.3.2	Application to real data: BMD inferences on a pilot EyeLink dataset	140

4.3.3	Application to real data: Dual Purkinje Image data . . . . .	142
4.3.4	Variants of BMD . . . . .	144
4.3.5	Application to real data: ADHD and Controls EyeLink dataset . . .	149
4.4	Discussion . . . . .	152
4.4.1	Caveats of the BMD algorithm . . . . .	153
4.4.2	Conceptual extensions of the BMD algorithm . . . . .	154
4.4.3	Microsaccades and perceptual variability in ADHD . . . . .	155
4.5	Supplementary . . . . .	156
4.5.1	Mathematical details of the BMD algorithm . . . . .	156
4.5.2	Parameter inference in simulated and DPI data . . . . .	166
4.5.3	Microsaccade kinematics in the EyeLink pilot data . . . . .	167
4.5.4	Microsaccades inferred in the EyeLink pilot data . . . . .	168
4.5.5	BMD caveat: inference on low-noise simulated data . . . . .	169
<b>5</b>	<b>Discussion</b>	<b>170</b>
	<b>References</b>	<b>176</b>

# List of Figures

2.1	Schematic of laboratory visual search tasks . . . . .	16
2.2	Tasks: Localization and Detection, each with conditions Perception and Memory. . . . .	24
2.3	Stimuli spacing in Experiment 1 vs Experiment 2. . . . .	24
2.4	Generative model for Localization and Detection . . . . .	27
2.5	Experiment 1 Localization : proportion correct . . . . .	37
2.6	Experiment 1 Localization: responses . . . . .	37
2.7	Experiment 1 Localization: proportion correct as a function of distractor heterogeneity . . . . .	38
2.8	Experiment 1 Localization: proportion correct and model fits . . . . .	39
2.9	Experiment 1 Localization: responses and model fits . . . . .	39
2.10	Experiment 1 Localization: model comparison . . . . .	40
2.11	Experiment 1 Localization parameters . . . . .	41
2.12	Experiment 1 Localization: reaction times . . . . .	41
2.13	Experiment 1 Detection: optimal-observer model NoLapses . . . . .	42
2.14	Experiment 1 Detection: model comparison . . . . .	43
2.15	Experiment 1 Detection parameters . . . . .	44
2.16	Experiment 1 Joint fits Localization and Detection . . . . .	45

2.17	Experiment 1 Joint parameters . . . . .	46
2.18	Experiment 2: proportion correct and model fits . . . . .	48
2.19	Experiment 2: responses and model fits . . . . .	48
2.20	Experiment 2 Localization: distractor heterogeneity, data and model . . . . .	49
2.21	Experiment 2 Localization: model comparison . . . . .	50
2.22	Experiment 2 Localization parameters . . . . .	51
2.23	Experiment 2 Localization: reaction times . . . . .	51
2.24	Experiment 2 Detection: optimal-observer model NoLapses . . . . .	52
2.25	Experiment 2 Detection: model comparison . . . . .	53
2.26	Experiment 2 Detection parameters . . . . .	53
2.27	Experiment 2 Joint fits Localization and Detection . . . . .	55
2.28	Experiment 2 Joint parameters . . . . .	56
2.29	Quantile binning . . . . .	61
3.1	Task design . . . . .	67
3.2	Dissociation of perceptual and executive processes . . . . .	69
3.3	TIMO and RT . . . . .	78
3.4	Psychometric curve parameters . . . . .	83
3.5	Classification of participants . . . . .	87
3.6	Training information . . . . .	93
3.7	Further characterization of responses . . . . .	95
3.8	exGaussian parameters fits to reaction times . . . . .	96

3.9	Model comparison: reaction times . . . . .	97
3.10	RT iqr . . . . .	98
3.11	Distribution of stimuli . . . . .	99
3.12	Model comparison: psychometric curves . . . . .	100
3.13	Psychometric curves and parameters for both models . . . . .	102
3.14	Effect of learning . . . . .	103
3.15	Effect of eye tracking . . . . .	104
3.16	Correlations across metrics . . . . .	106
3.17	Correlations for each condition . . . . .	109
4.1	Microsaccades under different eye tracker noise levels . . . . .	119
4.2	Generative model of the eye position measurements time series . . . . .	121
4.3	Priors over durations and velocities . . . . .	123
4.4	Performance of BMD and EK6 on simulated data . . . . .	137
4.5	Performance of all algorithms on simulated data . . . . .	138
4.6	Performance of all algorithms on simulated data: ROC . . . . .	139
4.7	Real data example inferred microsaccades . . . . .	141
4.8	Performance of algorithms on DPI data plus noise . . . . .	143
4.9	Performance of algorithms on DPI data plus noise:ROC . . . . .	144
4.10	Performance of BMD variants . . . . .	146
4.11	Performance of BMD variants on simulated data: ROC . . . . .	147
4.12	Performance of BMD variants on DPI data: ROC . . . . .	147

4.13 Schematic of microsaccade detection algorithm . . . . .	148
4.14 Microsaccade rate across time within a trial . . . . .	149
4.15 Spearman correlations of microsaccade rate with behavior . . . . .	150
4.16 Main sequence of BMD and EK6 . . . . .	151
4.17 Solving integral A . . . . .	162
4.18 MCMC steps . . . . .	163
4.19 Parameter recovery in simulated data . . . . .	166
4.20 Microsaccade kinematics in EyeLink dataset 1 . . . . .	167
4.21 Inferred microsaccade rates in EyeLink dataset 1 . . . . .	168
4.22 BMD microsaccade inference is robust to priors . . . . .	168
4.23 BMD failures under low-noise . . . . .	169

# List of Tables

2.1	Optimal-observer models and parameters . . . . .	32
3.1	Correlations across metrics . . . . .	84
3.2	Demographics . . . . .	92
3.3	Psychiatric scores . . . . .	92
3.4	Correlations across psychiatric scores . . . . .	94
3.5	Group specific correlations . . . . .	107
3.6	Correlations by symptom type . . . . .	108
3.7	Logistic regression coefficients . . . . .	110
3.8	Linear regression coefficients . . . . .	111
4.1	BMD algorithm steps . . . . .	126
4.2	BMD parameter inference . . . . .	166

# Chapter 1

## Introduction

Every second, our sensory systems are inundated with a plethora of information. Our brains have the job of filtering out what is irrelevant or unchanged and attending to behaviorally relevant information, sometimes by directing our eyes towards it, and other times just by directing 'our mind's eye', or covert attention. This thesis builds on behavioral paradigms that aim to engage visual covert attention. In this introductory chapter, we provide a brief overview of visual attention, describe a few classic behavioral paradigms used to study it, as well as describe psychophysical measures used to characterize and quantify its influence on the precision of perceptual representations. Next, we provide background about attentional-deficit hyperactivity disorder (ADHD), disorder in which some aspects of attention are thought to be impaired, but the exact nature of the attentional impairments is an ongoing investigation. Last, we review the roles of microsaccades in perception and cognition, in the context of the specific attentional paradigms we de-

scribe.

## 1.1 Visual attention

Imagine you are driving a car on a road in your favorite national park. Even during a snippet of time when you are fixating your eyes on the center of the road, you are still covertly allocating attention and monitoring information in your visual periphery. Your attention can be *divided* across multiple items/locations – other cars on other lanes, potentially other people, and landscape elements such as trees. In several cases, the more items your attention is divided over, the less precise your perceptual representation of each one of them will be. While you are still fixating on the road, a sign pointing rightward could appear, akin to a spatial cue used in laboratory tasks of *selective attention*. Before moving your eyes and turning, you will likely covertly allocate your attention in that direction and focus your neural processing resources to sharpen the visual information in that particular region of space. In a sense, divided attention and selective attention hint at opposite processes, one identified by the cost of splitting processing resources across items and another by the benefit of increasing allocation of neural resources to behaviorally relevant items. Given that representing information with spikes is costly and our brains have limited energetic resources, it might be useful to allocate more neural resources and represent certain items with more precision than others (Carrasco, 2011, 2014; Lennie, 2003), for instance a cyclist relative to a car. Efficient and flexible allocation of costly brain resources can be accomplished through attentional mechanisms. First, we attempt to break down attentional processes into subcomponents and identify specific behavioral paradigms that probe them.

Divided attention and selective attention represent two main types of attention;

they have been studied in recent decades with two main classes of paradigms: visual search (Palmer et al., 1993; Treisman and Gelade, 1980) and respectively visual spatial orienting (Posner, 1980). An example of a visual search task is detecting a specific item (for instance, a specific letter) in a display of items. In visual search tasks, performance metrics (i.e., accuracy, reaction times) tend to degrade as the number of items increase, which has been called the *set size effect*. Under certain experimental conditions, attention can be said to be divided across the items in a display (Palmer, 1995, 2014). Attention can also be divided across features, such as in dual task paradigms, where in the dual condition both features of an item are relevant to the task. Tasks of visual spatial orienting usually entail the presentation of a line cue, either centrally or peripherally, followed - if the cue is valid - by the presentation of an item in the cued location. Next, we present more information on visual search and visuo-spatial orienting, as well as measures and models to probe attentional processes with these paradigms.

## 1.2 Visual search and the feature integration theory of attention

Perhaps the most well-known visual search work started with (Treisman and Gelade, 1980). In one condition called feature search, observers had to detect whether a specific letter was present in a display of letters, say a T. In another condition called conjunction search, observers were supposed to detect whether, say, a green T was present in a display of colored letters. (Treisman and Gelade, 1980) developed the feature integration theory (FIT) of attention to explain the result that reaction times were approximately constant with set size for feature searches (letter T), but increased with set size for conjunction searches (green letter T). FIT posits that in a first pre-attentive stage, all features of the items are processed in parallel - in the above example, colors and letters - and in a second

stage, selective attention is engaged serially to bind these loose features into joint objects, so that the observer can ultimately determine whether the target of interest (here green letter T) was present. In this FIT account, attention is not thought to be divided across all items, but to serially be allocated in a selective fashion from item to item.

### **1.3 Visual search, signal detection theory and divided attention**

While highly influential, the FIT of attention came under scrutiny on several grounds. A particular group of researchers, which applied signal detection theory (SDT) models to visual search (Cameron et al., 2004; Eckstein et al., 2000; Palmer et al., 1993, 2000) emphasized that key elements of FIT such as limited processing or serial allocation of attention are not needed to explain set size effects in more general visual search paradigms. Instead, in the SDT account, unlimited parallel processing of stimuli can happen even during conjunction search; the set size effect could be due to the fact that the more distractors there are, the higher the chance a noisy representation of one of them could be confused with the target (Palmer, 1995). Work in this framework used a few conceptual distinctions: unlimited versus limited processing capacity, parallel versus nonparallel processing and interactive versus noninteractive processing (Palmer, 2014). Unlimited processing entails independent processing of the stimuli, such that the addition of a new stimulus does not impair the overall quality of information processing. In parallel processing, the performance metric does not degrade with the size of the display. Lastly, processing is said to be non-interactive if one can assume that the processing of one item is not influenced by the other items in the display.

Within this line of research, the visual search task designs used were more likely to

engage divided attention. When observing the set size effect, we might be tempted to interpret it as a divided attention effect. However, it is not trivially true that attention is divided across the items in a search display - experimental designs have been optimized to maximize the chance that observers are indeed dividing their attention across items, as opposed to, for instance, suffering interference due to sensory effects or selectively allocating their attention sequentially from item to item (Palmer, 1995, 2014). Such experimental designs had brief stimulus presentation times and/or eye tracking to prevent observers from making sequential eye movements to different items, as well as to keep the stimuli in the search array at the same eccentricity and space them widely to prevent performance decrements due to low-level crowding phenomena (Palmer, 1995, 2014; Palmer et al., 1993).

Originally developed for radar detection, signal detection theory provides a quantification of component processes of decisions about stimuli in the presence of uncertainty or noise (Peterson et al., 1954). Its formalism has also proved useful in perceptual psychophysics (Green and Swets, 1966). Drawing on work applying SDT to simpler tasks such as one or two item detection or discrimination paradigms (see next section), researchers that applied SDT to visual search also used stimuli that are simple and can be readily parametrized (similarities among oriented bars are much easier to quantify than similarities among letters). According to SDT, stimuli are perceptually *encoded* in a noisy way by an observer, who eventually makes a *decision* about whether, say, the target was present.

Visual search is more difficult a task to explain with SDT than one or two items detection or discrimination paradigms; this difficulty comes from a larger freedom to specify a way to combine the information from the noisy encoded stimuli and ultimately compute

a decision rule (Palmer et al., 2000). Among these possible ways to combine information into a decision rule, optimal-observer models stand out by having a unique Bayesian decision rule, derived by taking into account the statistical distribution of the stimuli and making plausible explicit assumptions, for instance about task operations, and the noise distribution on encoding (Eckstein, 2011; Ma et al., 2011; Vincent, 2011). In the context of some optimal-observer models, the amount of noise with which items are encoded can be quantified with precision parameters (see Section 2.3.1, as well as (Ma et al., 2011; Mazzyar et al., 2012)). These behavioral precision parameters are thought to have a neural correspondent in the amplitude or gain of a neural population representing the stimuli (Ma et al., 2014).

## 1.4 Selective attention

In visual search tasks, selectively cue-ing spatial attention has been shown to attenuate the decay of performance with set size (Cameron et al., 2004). Like divided attention, selective attention can operate on space or features to enhance their perceptual processing. Selective spatial covert attention generally improves task performance in the cued location, and tends to hurt performance in the uncued location (for reviews, see (Carrasco, 2011, 2014)). It can be engaged either exogenously, via a peripheral cue or stimulus, or endogenously, in a slower, voluntary, goal driven way, for instance via a central cue (Müller and Findlay, 1988). Feature-based attention can enhance other feature dimensions of a stimulus, say its color, or within the same feature dimension, say color, a specific feature value, such as a specific shade of red. Spatial attention and feature-based attention have been shown to depend on distinct, but substantially overlapping fronto-parietal circuits (Greenberg et al., 2010; Scolari et al., 2014). Additionally, the effects of spatial covert atten-

tion also depend on subcortical activity, specifically the superior colliculus (i.e. (Krauzlis et al., 2013; Lovejoy and Krauzlis, 2017)), as well as the thalamus (i.e. (Halassa and Kastner, 2017; Rafal and Posner, 1987)).

It has been shown that covert selective attention quantitatively affects visual perception; these effects have been quantified with coarser task metrics such as accuracy and reaction times. Accuracy and reaction times can trade-off in some cases, limiting the experimenter's ability to make robust inferences about the effect of a manipulation, such as covert attention. Another disadvantage is that tasks that use reaction time as the dependent variable tend to have longer stimuli presentations, during which observers might move their eyes to the target. Additionally, an increase in reaction time is hard to interpret mechanistically in the absence of modeling since it can be due to slower processing in several possible task stages - attentional allocation, perceptual encoding, decision making, location of their decision threshold, motor preparation or response execution.

Signal detection theory has also been used to study the effects and mechanisms of selective attention. In this context, as opposed to visual search, researchers can draw on classical one or two items detection or discrimination psychophysical paradigms. While not entirely without trade-off concerns, SDT has been employed to separate sensitivity in *perceptual encoding* from *decision processes*. Stimulus strength is parametrically manipulated and a psychometric function is fitted to the observer's responses with stimulus strength, yielding the corresponding parameters perceptual sensitivity  $d'$  and response bias. Several other parametrizations of psychometric curves are also frequently used (Klein, 2001). Covert selective attention has been shown to modulate the parameters of psychometric curves; much of this work has examined contrast sensitivity psychometric functions. Interestingly, endogenous attention has been shown to improve performance by

increasing the contrast gain, while exogenous attention through increases in both contrast and response gain (Cameron et al., 2002; Carrasco, 2011; Ling and Carrasco, 2006).

Even without delving into the extensive literature from which the above are just a few examples, we see that attentional manipulations can be introduced in tasks as described above and their effects on the observer's precision of representation can be inferred with tools from signal detection theory. Applying such paradigms to the study of disorders in which some aspects of attention might be impaired, ADHD (Mullane and Klein, 2008; Roberts et al., 2017), autism (Grubb et al., 2013)) can help pinpoint deficits and hopefully ultimately inform interventions.

## **1.5 Attention-deficit/hyperactivity disorder (ADHD)**

Studying a disorder in which some aspects of attentional processes are thought to be impaired can provide both insight into the mechanism of the disease and thus opportunities for better diagnosis and/or intervention, as well as possibly new knowledge about aspects of attention in general.

Attention-deficit/hyperactivity disorder (ADHD) is a complex, chronic, and heterogeneous neurodevelopmental disorder (Diagnostic and Statistical Manual (DSM)-V, 2013). DSM-IV (1994) characterized three subtypes based on different combinations of inattentive symptoms and hyperactive-impulsive symptoms: predominantly inattentive, predominantly hyperactive impulsive and combined type (Willcutt, 2012). Although ADHD is traditionally thought of as a pediatric disorder, symptoms often persist into adulthood, affecting 4.4 % of adults in the United States (Kessler et al., 2006). While symptoms of hyperactivity and impulsivity decrease with age, symptoms of inattention and disorganiza-

tion tend to persist and may even become more taxing as adolescence and adulthood bring increasing educational, occupational and self-management demands (Biederman, 2000; Stevens et al., 2012; Wilens, 2004). Adults with ADHD are at higher risk of underemployment, and often suffer from comorbid psychiatric disorders such as anxiety or substance abuse (Biederman J, 1991; Kessler et al., 2006).

ADHD is highly prevalent throughout the world, ranging from 1-7% (Fayyad et al., 2007). Additionally, ADHD diagnosis rates show a large variability across the world (Polanczyk et al., 2007) and the US (Visser et al., 2014), raising concerns of either under or over diagnosis (Casal et al., 2018; Pierre, 2013). ADHD diagnosis is mainly assessed based on patients' self-reported answers to psychiatric questionnaires probing daily life function according to the DSM symptoms. Together with other factors, the subjectivity inherent in such diagnosis procedures could account to some extent for the variability in diagnosis.

Additionally, such psychiatric questionnaires provide insufficient insight into the potential differences in neural information processing that give rise to the ADHD symptoms. The daily life impairments captured by psychiatric diagnosis questionnaires can be broken down into component brain processes, processes which can be quantified with behavioral tasks and metrics, yielding "psychomarkers" or biomarkers. Such efforts are part of the emerging field of computational psychiatry (Montague et al., 2012; Redish and Gordon, 2016).

## 1.6 Microsaccades

Insight about the allocation of attention can come from eye movements. Our eyes incessantly scan the visual environment to acquire information and update our knowledge, as either we or objects move through the world. Eye movements such as large saccades and fixations bring objects onto the fovea for higher-resolution processing and have thus been used as measures of attention (Duc et al., 2008). But even when we attempt to fixate our gaze on a stationary object, our eyes are always in motion. At this smaller scale, fixational eye movements, microsaccades in particular, could compensate for non-homogeneous resolution within the fovea (Poletti et al., 2013) and provide information about the allocation of covert attention (Hafed and Clark, 2002). Behaviorally (Otero-Millan et al., 2013, 2008; Zuber et al., 1965) and neurophysiologically (Hafed and Krauzlis, 2012), saccades and microsaccades fall on a continuum of oculomotor behaviors (for a review, see (Martinez-Conde et al., 2013)); their generation (Hafed et al., 2009; Munoz and Wurtz, 1995) and suppression (Hafed and Krauzlis, 2010; Robinson and Wurtz, 1976) have been shown to be controlled by the superior colliculus (SC).

Fixational eye movements fall into three categories: drift, tremor, and microsaccades (Ciuffreda and Tannen, 1995). For the majority of the fixation time, the eye is in a drift motion state, which can be described as a random walk (Cornsweet, 1956; Ditchburn and Ginsborg, 1953; Ratliff and Riggs, 1950). Drift movements cover about a dozen retinal photoreceptors (Martinez-Conde and Macknik, 2004; Ratliff and Riggs, 1950) and have relatively low amplitude and velocity, whereas tremor is an oscillatory movement of high frequency and very small amplitude that rarely exceeds the size of one photoreceptor (Findlay, 1971). By contrast, microsaccades can carry the retinal image over a dozen

or more photoreceptors, depending on their amplitude (Rolfs, 2009). For instance, if drift movements have carried an image away from the area with the highest photoreceptor density (fovea), microsaccades can serve to quickly bring it back (Martinez-Conde and Macknik, 2004). Microsaccades occur a few times per second, as ballistic, higher velocity movements of larger amplitude that appear as a linear part of the eye movement trajectory (Engbert et al., 2011).

The high velocity and amplitude of microsaccades amongst fixational eye movements makes them candidates for distinct roles in perception and cognition. Insight into these roles can come from the modulation of microsaccade rate and direction across task periods. Microsaccades that occur before stimulus presentation can facilitate orientation discrimination of the stimuli presented in their direction relative to the ones presented opposite to it (180 deg), akin to shifts of covert spatial attention (Yuval-Greenberg et al., 2014). Crucially, this study employed the rarely used method of online detection of spontaneous microsaccades and targetted stimulus presentation either in their direction or opposite. Analyzing microsaccades during the stimulus period - if a stimulus is presented for a sufficiently long period of time, i.e. 1000 ms in (Poletti et al., 2013) - studies discovered that microsaccades can be useful in high-acuity tasks such as threading a needle (Ko et al., 2010; Poletti et al., 2013) or help prevent or counteract visual fading (Martinez-Conde et al., 2006; McCamy et al., 2014).

Around 100-200 ms after the onset of a stimulus or a cue, the rate of microsaccade is suppressed (Engbert and Kliegl, 2003; Hafed and Ignashchenkova, 2013; Rolfs et al., 2008). This fast inhibition might serve to facilitate encoding (Herrington et al., 2009; Martinez-Conde et al., 2013; Scholes et al., 2015). In line with this, observers can decrease their microsaccade rates voluntarily or when engaged in tasks that require high acuity

(Engbert, 2006). Early on, this inhibition was thought to be the result of the direct retina to SC pathway signaling a general detection of sensory input (Engbert, 2006); more recent accounts emphasize that this suppression might be modulated by higher-level cognition, for instance cortex-dependent conscious detection of stimuli (White and Rolfs, 2016), or perceptual decision formation (Loughnane et al., 2018). The directions of the few microsaccades that occurred during this suppression period were shown to follow the direction of the spatial cue (Pastukhov and Braun, 2010).

This suppression in microsaccade rate is followed by a rebound at around 300-350 ms, likely modulated by higher cognitive processes (Rolfs et al., 2008), and ultimately a return to baseline (Engbert and Kliegl, 2003; Martinez-Conde et al., 2009; Rolfs, 2009). Studies using the Posner spatial attention cueing paradigm - (Engbert and Kliegl, 2003) with an endogenous central cue and (Hafed and Clark, 2002) with a slightly peripheral cue - showed that the distribution of directions of the microsaccades that occurred during the rebound stage showed a tendency towards the direction of the spatial cue. Following studies confirmed this result for exogenous covert attention (Laubrock et al., 2005; Rolfs et al., 2004, 2005), and also for endogenous covert attention engaged with central cues (Gowen et al., 2007; Laubrock et al., 2007), but sometimes with weaker and slower effects (Laubrock et al., 2005). In light of these spatial cueing studies, microsaccades gained recognition in the literature as an index of covert attention (Engbert, 2006; Engbert and Kliegl, 2003; Rolfs, 2009; Yuval-Greenberg et al., 2014), not without some controversy (Collewijn and Kowler, 2008; Horowitz et al., 2007), the nuances of which have been examined in detail (Laubrock et al., 2010).

Even in the absence of attentional cues, the pattern of microsaccadic suppression followed by rebound - also called the microsaccade rate signature - has been shown to be

modulated by the stimuli features, especially contrast (Bonneh et al., 2015; Scholes et al., 2015). In particular, aspects of the microsaccade rate signature - inhibition magnitude, rebound magnitude - were able to predict human contrast sensitivity thresholds (Scholes et al., 2015). Lastly, in some studies the magnitude of microsaccade suppression has been observed to be higher with increased attentional load (Pastukhov and Braun, 2010) or in a task with increased working memory load (Siegenthaler et al., 2013). The microsaccade roles presented here are by no means exhaustive, but either pertain to the work in this thesis and/or provide a broad sense of their importance.

Microsaccades have been shown to have distinctive features in several neurological disorders such as Parkinsonian disorders (Otero-Millan et al., 2013), Alzheimer’s disease (Kapoula et al., 2013), progressive supranuclear palsy (Otero-Millan et al., 2011) (for a review see (Alexander et al., 2018)). A few recent studies extended such investigations into microsaccades in ADHD (Dankner et al., 2017; Fried et al., 2014; Panagiotidi et al., 2017; Roberts et al., 2017) (see Section 4.1.1).

## 1.7 Dissertation outline

The specific tasks, models and measures of visual attention employed in this thesis are outlined below.

In Chapter 2, we examine visual search with heterogenous distractors. We draw on visual search experimental designs that have been optimized to maximize the chance that observers are engaging divided attention and minimize the chance of selective attention or other confounds (Palmer, 1995, 2014). We characterize in detail the behavior of human observers in two tasks, localization and detection, in both what we call perception

and memory conditions. Observers have to attend to a particular orientation value, and either detect it or localize it within a search array, process which might require them to divide their attention across the oriented patches. The relevant feature value is presented either before or after the search array, yielding perception and memory conditions. We ask whether an optimal observer model with variable precision encoding can capture the behavioral data.

In Chapter 3, we exemplify the general approach in computational psychiatry of designing behavioral tasks to identify differences in brain processes in health and disease, here ADHD. To this end, we develop a task in which we aim to engage selective attention, specifically covert endogenous spatial attention, and additionally on half of the trials feature dimension switches. Participants also have to keep track of and act on 2 or respectively 4 stimulus-response rules. Stimuli are 2 ellipses, each with two features, orientation and color. We use psychometric curves to quantify perceptual variability across conditions. This task design segregated perceptual variability from response selection/cognitive control/executive function, which was measured as button presses for the wrong space or feature. Of note, while perceptual precision and executive function are segregated as two different task metrics, the attentional and executive demands of the task likely impact perceptual variability. Whereas tasks that studied perception in isolation did not find differences in ADHD (see (Fuermaier et al., 2017) for a review), it is possible that taxing simultaneously other processes might allow for differences in perceptual precision to emerge.

In Chapter 4, we attempt to develop a new measure of microsaccades. We got interested in microsaccades due to their potential as measures of covert attention, but we first focused on developing a detection algorithm. Specifically, we develop a generative model of fixational eye movement measurements, and based on this model, a Bayesian inference

algorithm for microsaccade detection, which we call BMD. We compared BMD against other algorithms, mainly the most widely used velocity threshold algorithm (Engbert and Kliegl, 2003), as well as a newer unsupervised clustering algorithm (Otero-Millan et al., 2014a). To be able to compare the performance of algorithms, specifically robustness to measurement noise, we needed to have a reasonable assumption for ground truth microsaccade detection - we took that to be the average of all algorithms' inferences on a very low-noise dataset measured with a DPI instrument, made available to us by (Cherici et al., 2012). Lastly, we applied the BMD algorithm to the fixational eye movement time series in Chapter 3 and made a preliminary attempt to uncover an oculomotor mechanism underlying perceptual variability.

## Chapter 2

# Visual search with heterogenous distractors

*...perceptions received by the ears or by reflection can be most easily retained in the mind if they are also conveyed to our minds by the mediation of the eyes, ... so that we keep hold of as it were by an act of sight things that we can scarcely embrace by an act of thought....But these forms and bodies, unlike all the things that come under our view, require an abode, inasmuch as a material object without a locality is inconceivable. Consequently ... one must employ a large number of localities which must be clear and defined and at moderate intervals apart, and images that are effective and sharply outlined and distinctive.*

Cicero, de Oratore, 55 BC

---

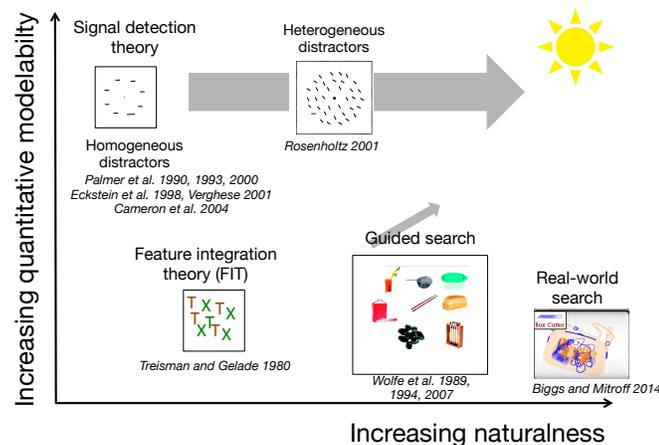
## 2.1 Introduction

Common real-life examples of visual search include detecting a friend in a crowd, locating one's keys in a tray of random items, and detecting a weapon during airport security screening (see (Biggs and Mitroff, 2014) for a review of the challenges in visual search for security screening). Using such complex images in behavioral tasks in the laboratory would render the studies naturalistic; however natural images have a high degree of complexity and are thus very hard to model. "Modelability" is desirable since in addition to using stimuli that conserve some of the properties of natural images, we also want to be able to capture the data with simple quantitative models, which can help dissociate component processes, such as sensory encoding and decision processes.

We can depict laboratory visual search tasks as points situated in a two dimensional plane of naturalness and quantitative modelability, where we ideally want to reach the maximum of each dimension (Figure 2.1). Textbook visual search entails the feature integration theory of attention (FIT) (Treisman and Gelade, 1980), line of work which utilized tasks such as detecting whether there is a green T among a search array of colored letters. Such tasks are not very naturalistic and are also hard to model quantitatively. An extension of FIT was made by Wolfe and colleagues who developed the guided search model, which is more quantitative; updated versions have gotten even more so over the years (Wolfe, 2007). However, it still relies on a lot of built-in complexity and it has very many parameters, which makes interpretability difficult.

Alongside these approaches to visual search, there has been a specific subgroup of researchers studying visual search in the signal detection theory (SDT) framework (Pal, 1990; Cameron et al., 2004; Eckstein et al., 2000; Palmer et al., 1993, 2000). To make lab-

oratory visual search more tractable to parametrize and model, it has been common to make the distractors *homogeneous* – identical to each other (Nolte and Jaarsma, 1967; Palmer et al., 2000; Peterson et al., 1954; Vergheese, 2001). In this example, the task is to detect whether there is a longer line among short lines, where all the short lines have the same length. When distractors are homogeneous, the difference between the target and distractors can be captured by a single number. Signal detection theory has been used to capture how performance degrades as a function of the number of distractors in tasks of target detection, localization and identification (Cameron et al., 2004; Eckstein et al., 2000; Palmer et al., 2000). One way to increase naturalness is to use heterogeneous distractors. Preliminary efforts have been made to extend this framework to tasks with heterogeneous distractors (Rosenholtz, 2001), but more work is needed for more thorough characterization.



**Figure 2.1** Schematic of laboratory visual search tasks projected on a two dimensional space of naturalness and quantitative modelability.

Even with simple stimuli, visual search is not straightforward to model quantitatively. As we mentioned in the introduction, this is because visual search is a more complex fam-

ily of tasks than the single-item detection and discrimination. This additional complexity arises partly because information must be integrated across items, which raises the question of what rule the brain uses for this integration (for example the max rule (Nolte and Jaarsma, 1967; Verghese, 2001), the sum rule (Baldassi and Verghese, 2002; Graham et al., 1987; Palmer et al., 2000), or a Bayesian rule (Ma et al., 2011, 2015; Shen and Ma, 2017; Vincent et al., 2009)).

While the choice of homogenous distractors has the advantage of making the calculations tractable, it is not representative of many situations of real-world search in which distractors are heterogeneous. Therefore, it is important to also study heterogeneous visual search; we will do so while still maintaining strict psychophysical control and the ability to model the task using a concise mathematical process model. Although much smaller than the literature on homogeneous visual search, there is substantial past work on heterogeneous visual search. (Duncan and Humphreys, 1989) showed and proposed that search efficiency decreases as the similarity between targets and distractors increases, and also as the similarity between non-targets decreases. Their stimuli were letters and their independent variables were accuracy and reaction times, thus leaving room for improvement in terms of quantitative modelability.

A first approach to model visual search with heterogenous distractors was made by (Rosenholtz, 2001), who used a two-interval forced choice task and developed an extension of signal detection theory approaches previously applied to the study of homogeneous distractors; one problem was that the number of subjects was low. Additionally, there are several ways to choose heterogenous distractors and conclusions might vary with these choices. (Ma et al., 2011) varied reliability and found that an ideal-observer model captured target detection behavior for both homogenous distractors and heterogeneous

distractors drawn from a uniform distribution; however, this work did not vary set size. (Mazyar et al., 2012) used heterogeneous distractors drawn from a uniform distribution and varied set size in a target detection task. An ideal-observer model only fitted well if the additional allowance was made that the mean encoding precision of a given item varied with set size. This was true regardless of whether the identity of the target was revealed before or after the search array.

In spite of these efforts, there is much more to be done to understand heterogeneous visual search from a modeling perspective. Here, we attempt to do so in a systematic manner, restricting ourselves to orientation search with distractors drawn independently from a uniform distribution. One interesting aspect of uniformly distributed heterogeneous distractors is that by chance they might be more or less confusable with the target; this offers the opportunity to plot (and model) a new kind of psychometric curve, with fine grained target-distractor similarity, or confusability, as the independent variable.

Even within our restrictions, we can identify multiple factors that could affect search performance and mechanisms:

- Set size: as set size increases, performance usually decrease in searches with heterogeneous distractors, and to explain this, it seems to be necessary that encoding precision per item decreases with increasing set size (Mazyar et al., 2012, 2013).
- Task: while most visual search work uses detection or categorization, lots of real-world search is localization. For example, you know for a fact that your friend is in the crowd, but you need to localize her. Detection and localization have been studied jointly with homogenous distractors (Cameron et al., 2004; Liu, 2003), but not with heterogeneous.

- Time: One can search for an object by looking at the environment (which we will call *perception-based search*), or by recalling the visual environment from memory (which we will call *memory-based search*) (Mazyar et al., 2012).
- Space: Patterns of effects for set size, task and time might be dependent on the usage of a display widely spaced stimuli. Even apart from crowding (Bouma, 1970), it is conceivable that stimulus spacing affects behavior, especially in localization (Bays, 2016).

### 2.1.1 Structure of this chapter

The entire chapter is about orientation search with independent, uniformly distributed, heterogeneous distractors and brief presentation times. We report the results of two experiments: Experiment 1, in which stimulus spacing was large (60 deg), and Experiment 2, in which stimulus spacing was small (30 deg). In Section 2.2, we describe the paradigm of the Experiments 1 and 2. We systematically vary the four factors above. In addition, we vary the first three factors *within subjects*. The task was localization or detection in different blocks. To explore the time dimension, the search was based on perception (by having the target orientation revealed before the search array) or short-term memory (by having the target orientation revealed after the search array). In Section 2.3.1 we describe the general structure of the ideal-observer process model. We analyze the results by first restricting ourselves to localization in Section 2.4.1, examining the effects of set size and time. The effect of time amounts to a comparison between visual search based on perception and memory-based visual search. We will analyze all effects both in a model-free and in a model-based way; modeling serves to identify potential loci of any differences between conditions. We examine detection separately in Section 2.4.2 and then in Sec-

tion 2.4.3 we jointly examine detection and localization (within subjects); to our knowledge, we are the first to do so for heterogeneous distractors within a common modeling framework. In Section 2.5, we redo the analyses of previous sections for Experiment 2 (small spacing). In Section 2.6, we present the conclusions of these experiments and discuss limitations and future directions.

## 2.2 Experimental methods

### 2.2.1 Tasks and experiments overview

In Experiment 1, participants performed a target Localization task and a target Detection task, each with a Perception and a Memory condition. Across all 4 sessions, participants completed 800 trials in each condition. In the Perception conditions, subjects were presented briefly with a centrally presented target Gabor, followed by a delay and then visual search display of  $N$  Gabors, with set size  $N = 2, 3, 4, 6$ . The search display stimuli were placed adjacently at 60 deg angular intervals on an imaginary circle centered at fixation with radius (eccentricity) 5 degrees of visual angle (dva). In the Memory condition, the target search display was presented first, followed by the delay, and the centrally located target Gabor. Both Perception and Memory conditions require observers to engage memory to some extent, but while in the Perception condition observers only have to remember the orientation of one Gabor over the delay period, in the Memory conditions observers have to remember all  $N$  Gabors over the delay period.

In target Localization participants had to click on the location in the display that matched the orientation of the target Gabor. In target Detection, participants had to in-

dicating whether the orientation of the target Gabor matched any one of the orientations in the display. In target Detection, the target orientation  $s_T$  was present in the search display 50% of the time. Target and distractor orientations were independently drawn from uniform distributions on  $[-90, 90)$  deg.

Experiment 2 was almost identical to experiment 1 with just one key difference: the search display stimuli were placed adjacently at 30 deg angular intervals.

## 2.2.2 Experiments

Here we provide the methodological details of experiments 1 and 2.

### 2.2.2.1 Experiment 1

**Participants.** 11 participants (9 female, 2 male) performed the task upon providing informed consent. The study adhered to the Declaration of Helsinki and was approved by the NYU Institutional Review Board.

**Apparatus.** We displayed stimuli on a Dell 1907FPC LCD monitor with resolution  $1280 \times 960$  pixels and width 19" (with 16.1"/ 40.89 cm viewable) and 75 Hz refresh rate. A Windows computer running Matlab 8.2 (MathWorks, Massachusetts, USA) with the Psychtoolbox (Brainard,1997;Pelli,1997;Kleiner et al.,2007) displayed the stimuli. Subjects were located at approximately 60 cm from the screen. The screen background was mid-level gray.

**Stimuli.** On each trial, target and distractor orientations were drawn independently from a uniform distribution on  $(-90, 90)$  deg. Each stimulus was an oriented Gabor patch of

spatial frequency 2.85 cycles/deg, 80 % contrast, with standard deviation 0.26 deg subtending an image size of height and width both equal to 1.3 deg. Participants viewed searched displays in order to detect or locate the item that matched the orientation of the centrally presented target stimulus. The number of stimuli in the search display was 2, 3, 4 or 6, randomly interleaved. Since the search display stimuli were placed adjacently at 60 deg angular intervals on an imaginary circle and eccentricity was 5 dva, the center-to-center distance between two adjacent Gabors was 5 dva.

**Instructions.** Participants were informed in advance of the details of the experiment. In particular, we emphasized to them that while in Detection there is a 0.5 probability that the target is present in the search display, in Localization the target is always present in the search display.

**Localization.** Each trial in the Localization-Perception condition began with subjects fixating on a central fixation cross (diameter 0.15 dva) for 500 ms, followed by a centrally located target Gabor for 100 ms, a blank screen for 1000 ms, and then a search display for 100 ms (Figure 3.1). The short stimuli presentation times served to minimize the effect of eye movements. In Localization, the target was always present in the search display. The location of the target in the display was chosen randomly from the stimulus locations. Once the response screen appeared, participants had to use the mouse to click on the location where the stimulus orientation matched the target orientation. The possible location options were  $N$  white circles corresponding to the original locations of the stimuli. Thus, the Localization task was n-alternative forced choice (n-AFC) and the probability of responding correctly by chance was  $\frac{1}{N}$ . The Localization-Memory condition was identical to the Localization-Perception condition except that the temporal order of the target and the search display was reversed.

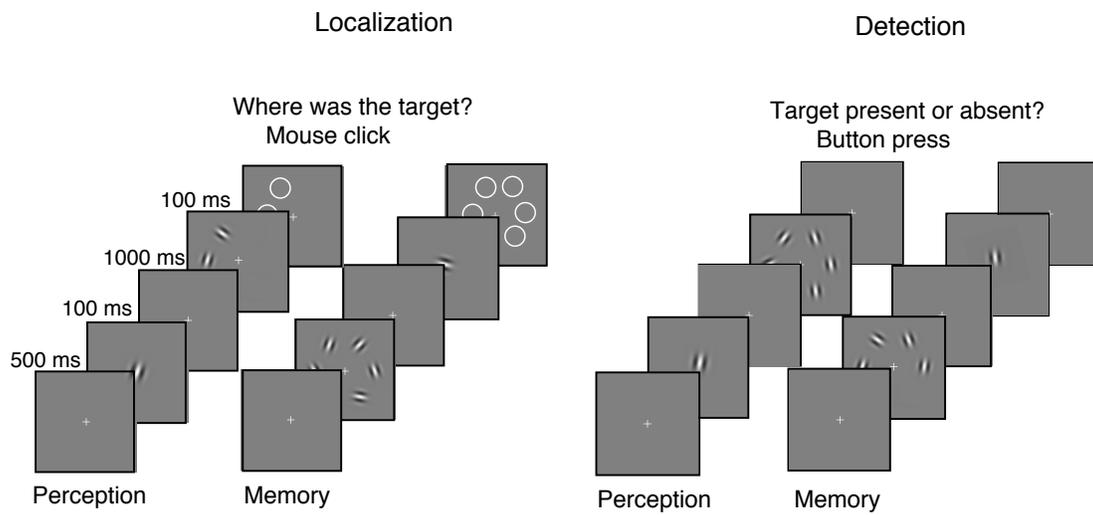
**Detection.** While in Localization the target was always present in the search display, in Detection the target was present in the display in half of the trials. On a target-present trial, the orientation of one of the items in the search display matched the orientation of the target Gabor. Subjects reported whether they believed the target orientation was matched or not with the orientation of an item in the search display by pressing one of two possible keys. Thus, the probability of responding correctly by chance in Detection was always  $\frac{1}{2}$ . The trials in Detection-Perception and Detection-Memory conditions followed the same sequence as in the analogous Localization conditions.

**Blocks and sessions.** The full experiment consisted of four sessions of 800 trials each. Each session was divided into 8 blocks after which the participants were presented with their percent correct scores so far grouped by condition. Conditions were labeled 1,2,3,4 for Detection-Perception, Localization-Perception, and respectively Detection-Memory and Localization-Memory. The order of the blocks was of the type 3-2-1-4-4-1-2-3, chosen randomly for each participant, but maintained across the sessions for that participant. Before the first session, participants performed a short training block consisting of 40 trials, 10 of each condition.

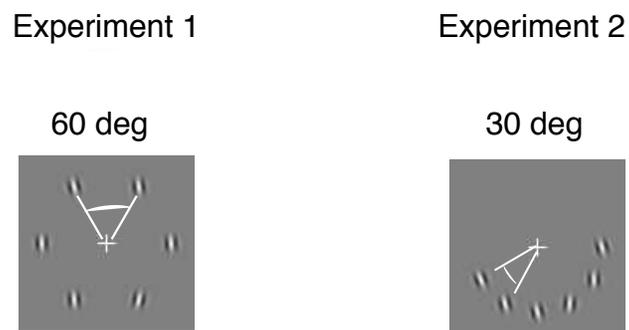
#### *2.2.2.2 Experiment 2*

Experiment 2 was almost identical to experiment 1 with just one key difference: the search display stimuli were placed adjacently at 30 deg angular intervals, and thus two adjacent Gabors were spaced at 2.6 dva center-to-center distance. The majority of the methods we described for Experiment 1 extend to Experiment 2, with the differences listed below.

**Participants.** 7 participants (4 female - one author, 3 male) performed the task upon



**Figure 2.2** Task design for Experiment 1. Trial sequence for Localization and Detection.



**Figure 2.3** The spacing of the stimuli in the search array in the Experiment 1 vs Experiment 2, depicted here for the highest set size 6.

providing informed consent. The study adhered to the Declaration of Helsinki and was approved by the NYU Institutional Review Board.

**Apparatus.** In Experiment 2, we displayed stimuli on a ViewSonic VX2475Smhl-4K LED monitor with resolution  $3840 \times 2160$  pixels and width 24" (with 23.1/ 59 viewable) and 60 Hz refresh rate. For this monitor, the screen background was also mid-level gray.

**Data and code.** Experiment, analysis and modelling code will be made available upon completion of this work at: [https://github.com/lianaan/Vis\\_Search](https://github.com/lianaan/Vis_Search).

## 2.3 Models

### 2.3.1 Optimal-observer models

Optimal-observer models attempt to simulate the steps that an observer would take to solve the task at hand by taking into account knowledge of the environment (in such tasks, the probability distributions of the stimuli), as well as prescribe that noise can corrupt the observer's performance at some point(s) during this step-wise process. At a minimum, optimal-observer models provide a starting point for modelling. Human observers have been shown to behave according to optimal-observer models in several simple perceptual tasks (Geisler, 2004; Ma et al., 2014). This might be the case as our sensory and decision-making systems might have been optimized throughout evolution for certain ecologically relevant tasks. Here, we choose to extend the investigation of this optimal-observer framework to include tasks of localization.

### 2.3.1.1 Step 1. Generative model: encoding

The generative models for target Localization and target Detection are presented in Figure 2.4. Optimal-observer theory for target Localization and Detection in visual search with heterogeneous distractors drawn from a uniform distribution has been previously derived (Ma et al., 2011; Mazyar et al., 2012). Here, we briefly present it.

We denote the location of the target with  $L$ , taking one of the values  $1, 2, \dots, N$ , where  $N$  is the set size (2, 3, 4 or 6), with any location being equally likely:  $p(L) = \frac{1}{N}$ . Since in target localization the target is always present at some location  $L$ , the  $\mathbf{T}$  array formally denotes that the target is present at location  $L$  (1) and absent at all other locations (0), which is  $\mathbf{T} = \mathbf{e}_L$ , where  $\mathbf{e}_L$  is the unit vector in  $\mathbb{R}^N$ ,  $\mathbf{e}_L = \underbrace{(0, \dots, 1, \dots, 0)}_{\text{only a 1 at position } L}$ .

In target Detection the target presence variable  $C$  can take values 0 or 1, with  $p(C = 0) = p(C = 1) = 0.5$ , and  $L = 0$  when  $C = 0$  and  $L$  same as above when  $C = 1$ .  $\mathbf{T}$  can be written as:

$$\mathbf{T} = \begin{cases} \underbrace{(0, \dots, 0)}_N, & \text{if } C = 0 \\ \mathbf{e}_L, & \text{if } C = 1 \end{cases}$$

In both generative models, each orientation stimulus  $s_i$  (with  $i = 1, 2, \dots, N$ ) follows a uniform distribution  $p(s_i) \sim U[-\frac{\pi}{2}, \frac{\pi}{2}]$ .

So far, all the aspects of the generative model were dictated by the the experimental design. Now, we assume additionally that the stimuli  $s_i$  are independently encoded as the noisy observations  $x_i$  and thus:

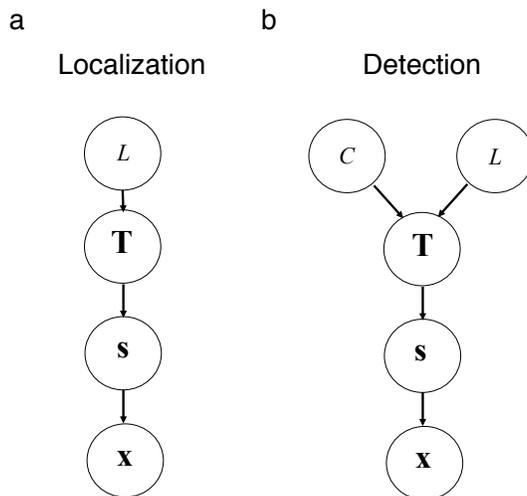
$$p(\mathbf{x}|\mathbf{s}) = \prod_i p(x_i|s_i)$$

In particular, we assume the noise follows a von Mises distribution. Since the support of the von Mises distribution is  $[-\pi, \pi)$ , we remap all orientations  $s_i$  from  $[-\frac{\pi}{2}, \frac{\pi}{2})$  to  $[-\pi, \pi)$  in our models and analyses. Thus, we can write:

$$p(x_i | s_i) = \frac{1}{2\pi I_0(\kappa_i)} e^{\kappa_i \cos(x_i - s_i)} \quad (2.1)$$

, where  $I_0$  is the Bessel function of the first kind of order 0. The inverse of the concentration parameter  $\kappa_i$  is related to the noise on item  $i$ . Resource in this framework is conceptualized as the Fisher information and is related to  $\kappa_i$  through:

$$J_i = \kappa_i \frac{I_1(\kappa_i)}{I_0(\kappa_i)} \quad (2.2)$$



**Figure 2.4** Generative models for (a) Localization and (b) Detection.

Additional assumptions in the variable precision encoding model (Fougnie et al., 2012; van den Berg et al., 2012) are about how the precision  $J$  varies across trials and items,

specifically according to a Gamma distribution with shape  $\frac{\bar{J}}{\tau}$  and scale  $\tau$ , and therefore expected mean  $\bar{J}$ . The relationship above between precision  $J$  and noise  $k$  is a monotonic function, which is easily invertible.

### 2.3.1.2 Step 2. Decision

The decision rules computed during the inference process are presented in detail in (Ma et al., 2011). We present them briefly here.

In the inference process, the optimal observer inverts the generative model and computes  $p(L|\mathbf{x})$ , with  $i = 1, 2 \dots N$  in Localization and ultimately  $p(C = 1|\mathbf{x})$  in Detection. The observer reports the location  $L$  for which  $p(L|\mathbf{x})$  has the highest value,  $\text{argmax}_i d_i$  and target present if  $p(C = 1|\mathbf{x}) \geq 0.5$ , which is equivalent to reporting target present when the global log posterior ratio, or decision variable  $d > 0$ .

In Localization, we derive the posterior as follows, making use of the facts that  $p(s_i)$  is uniform on  $[-\pi, \pi)$ ,  $p(s_i) = \frac{1}{2\pi}$  and  $\int_{-\pi}^{\pi} p(x_i|s_i)ds_i = 1$ .

$$\begin{aligned}
p(L|\mathbf{x}) &\propto p(L) \cdot p(\mathbf{x}|L) \propto p(\mathbf{x}|L) \\
p(\mathbf{x}|L) &= \int p(\mathbf{x}|\mathbf{s}) \cdot p(\mathbf{s}|L)ds \\
&= p(x_L|s_T) \cdot p(s_T) \cdot \prod_{i \neq L} \int p(x_i|s_i)p(s_i)ds_i \\
&\propto p(x_L|s_T) \frac{\prod_i \int p(x_i|s_i)ds_i}{\int p(x_L|s_L)p(s_L)ds_L} \\
&= \frac{p(x_L|s_T)}{\int p(x_L|s_L)p(s_L)ds_L} = e^{d_L} \tag{2.3}
\end{aligned}$$

This is the case since  $d_L$  represents the ratio of the probability of  $x_L$  coming from the target relative to the probability of it coming from a distractor.

In Detection, the global log posterior ratio takes the form:

$$d = \log \frac{p(C = 1|\mathbf{x})}{p(C = 0|\mathbf{x})}$$

We denote the participant's prior over the probability of target presence  $p(C = 1) = p_{\text{present}}$  and write  $d$  in terms of the prior and the log likelihood ratio:

$$d = \log \frac{p(\mathbf{x}|C = 1)}{p(\mathbf{x}|C = 0)} + \log \frac{p_{\text{present}}}{1 - p_{\text{present}}}$$

If each location is equally likely to contain the target,  $d$  can be written in terms of local log likelihood ratios  $d_i$  (Ma et al., 2011):

$$d = \log \frac{1}{N} \sum_{i=1}^N e^{d_i} + \log \frac{p_{\text{present}}}{1 - p_{\text{present}}}$$

where

$$d_i = \log \frac{p(x_i|T_i = 1)}{p(x_i|T_i = 0)}$$

Marginalizing over the distractors  $s_i$  and plugging in Equation (2.1),  $d_i$  becomes:

$$\begin{aligned} d_i &= \log \frac{\int p(x_i|s_i)p(s_i|T_i = 1)ds_i}{\int p(x_i|s_i)p(s_i|T_i = 0)ds_i} \\ &= \log \frac{p(x_i|s_T)}{\int p(x_i|s_i)p(s_i)ds_i} \end{aligned}$$

Again, since  $p(s_i)$  is uniform on  $[-\pi, \pi)$ ,  $p(s_i) = \frac{1}{2\pi}$  and because  $\int_{-\pi}^{\pi} p(x_i|s_i)ds_i = 1$ , we get:

$$\begin{aligned} d_i &= \log \frac{\frac{1}{2\pi I_0(\kappa_i)} e^{\kappa_i \cos(x_i - s_T)}}{\frac{1}{2\pi}} \\ &= -\log I_0(\kappa_i) + \kappa_i \cos(x_i - s_T) \end{aligned}$$

In Localization, since the priors over each possible location are equal, the optimal observer only has to compute the local log likelihood ratios  $d_i$  and report the location associated with the maximum  $\operatorname{argmax}_i d_i$ .

Some studies considered the possibility that the decision variable  $d$  is corrupted by noise. Since our detection task is an extension of the heterogenous task in (Mazyar et al., 2012), and in light of the results of (Shen and Ma, 2017) who did not find a contribution of decision noise to behavioral variability in that dataset, we do not include noise at the decision stage in our model.

### 2.3.1.3 Step 3. Model predictions

While the generative model, assuming parameters  $\theta$ , allows us to directly predict the observer's responses  $p(\hat{L}|\mathbf{x}, \theta)$  or respectively  $p(\hat{C}|\mathbf{x}, \theta)$  conditioned on the measurements  $\mathbf{x}$ , on a given trial we only have access to the stimuli  $\mathbf{s}$ , and thus we need to compute  $p(\hat{L}|\mathbf{s}, \theta)$ . To do this, we would have to marginalize over the measurements  $\mathbf{x}$  :

$$p(\hat{L}|\mathbf{s}) = \int p(\hat{L}|\mathbf{x}) \cdot p(\mathbf{x}|\mathbf{s})d\mathbf{x}$$

To estimate this integral, we took samples from this posterior distribution by simulating several (here 2000) measurements  $\mathbf{x}$  from  $\mathbf{s}$  and averaged over the corresponding outcomes  $\hat{L}$ . In models with lapses, the predictions were adjusted on every trial with the corresponding lapse  $\lambda$ :

$$p_{\text{lapse}}(\hat{L}|\mathbf{s}) = (1 - \lambda)p(\hat{L}|\mathbf{s}) + \lambda p_{\text{chance}}$$

where  $\frac{1}{N}$  for localization, with  $N$  being the set size and  $p_{\text{chance}} = \frac{1}{2}$  for detection.

#### 2.3.1.4 *NoLocLapse and NoDetLapse models*

For Localization, our main model, NoLocLapse (NL) has 5 parameters, 4 set size specific mean precisions  $\bar{J}$  and a  $\tau$  of the Gamma distributions. The main model for Detection, NoDetLapse (NL) has the same parameters as well as a bias for  $p_{\text{present}}$ .

#### 2.3.1.5 *Alternative models*

While the variable precision framework can in theory incorporate what look like lapses as trials with very low precision, there could be more systematic lapsing mechanisms.

Lapses are non-specific, they could capture any phenomena that leads the observer to respond independently of the stimulus presented. This could include non-target reports or swaps. Since non-target report frequencies increase with set size (Bays, 2016), if non-target reports drive a significant proportion of lapses, then a model in which localization lapses are allowed to increase with set size should capture the data better.

We explore two nested models with lapse rates. In LocLapse1 (L1), we allow for one lapse parameter, that we assume is the same across all the set sizes. In LocLapse2 (L2), we al-

low for two lapse parameters, that parametrize a linear increase of lapses with set size. These models and their parameters, as well as the corresponding ones for detection are listed in Table 2.1.

Parameters	Main parameters	Lapses				Total
		N=2	N=3	N=4	N=6	
NoLocLapse (NL)						5
LocLapse1 (L1)	$\bar{J}_2, \bar{J}_3, \bar{J}_4, \bar{J}_6, \tau$	$\lambda_L$	$\lambda_L$	$\lambda_L$	$\lambda_L$	6
LocLapse2 (L2)		$\lambda_{L2}$	$0.75\lambda_{L2} + 0.25\lambda_{L6}$	$0.5\lambda_{L2} + 0.5\lambda_{L6}$	$\lambda_{L6}$	7
NoDetLapse (ND)						6
DetLapse1 (D1)	$\bar{J}_2, \bar{J}_3, \bar{J}_4, \bar{J}_6, \tau, p_{\text{present}}$	$\lambda_D$	$\lambda_D$	$\lambda_D$	$\lambda_D$	7
DetLapse2 (D2)		$\lambda_{D2}$	$0.75\lambda_{D2} + 0.25\lambda_{D6}$	$0.5\lambda_{D2} + 0.5\lambda_{D6}$	$\lambda_{D6}$	8

**Table 2.1 Optimal-observer models and parameters.** We note the similarity of the models for detection and localization.

### 2.3.2 Modeling methods

We performed maximum-likelihood estimation (MLE) of the parameters in the models described above.

#### 2.3.2.1 Model fitting

For a particular model, the likelihood of a set of parameters  $\theta$  is the probability of the data given those parameters,  $p(\text{data}|\theta)$ . Log likelihood is denoted by LL. We assumed that trials are independent of each other and thus we could sum the log likelihoods across all trials:

$$\begin{aligned}
\text{LL}(\theta) &= \log p(\text{data}|\theta) \\
&= \log \left( \prod_{j=1}^{N_{\text{trials}}} p(\hat{L}_j | \mathbf{s}_j, \theta) \right) \\
&= \sum_{j=1}^{N_{\text{trials}}} \log p(\hat{L}_j | \mathbf{s}_j, \theta)
\end{aligned}$$

With the same rationale as described in "Model predictions", we approximated the Loglikelihood as the average of 2000 samples, which has a standard deviation of approximately 1 across several function evaluations.

To find the parameters  $\theta$  that maximize  $\text{LL}(\theta)$  we used an optimization method called Bayesian adaptive direct search (BADS) (Acerbi and Ma, 2017). For each dataset and model, we ran BADS with 20 starting points and chose the best fitting parameters among those. BADS is a hybrid of Bayesian optimization, performed with a local Gaussian process surrogate, and mesh adaptive direct search. Within BADS, we activated the noise setting options. `UncertaintyHandling = 1` and set the estimated noise size as options. `NoiseSize=1`. While we used 2000 samples and thus had a noise size of about 1, (Acerbi and Ma, 2017) showed good performance on a subset of our data with only 800 samples and a standard deviation of the Loglikelihood of approximately 3.5, updated accordingly in options. `NoiseSize`. (Acerbi and Ma, 2017) showed more generally and in particular for a subset of our data that more widely used optimization functions such as Matlab's `fmincon` or `fminsearch` are substantially worse at navigating noisy landscapes to find the global optima; `cmaes` with the noisy option was mildly worse.

We search the parameter space in the log range of some of the larger variables to en-

sure higher proximity of values on each parameter dimension and thus an easier problem for the optimization algorithm. The plausible parameter ranges were  $[\log(1.1), \log(300)]$  for the mean of the Gamma distribution  $\log(\bar{J})$  from which the precision  $J$  is drawn,  $[\log(10), \log(300)]$  for the scale parameter of the Gamma distribution  $\tau$ ,  $[0.3, 0.7]$  for  $p_{\text{present}}$  and  $[0, 0.99]$  for lapses  $\lambda$ .

### 2.3.2.2 Model comparison

Models with larger numbers of parameters lead to higher maximum log-likelihood (denoted  $LL^*$ ), but in order to find out if they actually fit better generally and do not just overfit the particular dataset, we calculated the model comparison metrics Akaike Information criterion, corrected for the number of trials ( $n_{\text{trials}}$ ), AICc (Hurvich and Tsai, 1989) and the Bayesian information criterion (BIC) (Schwarz, 1978). These metrics penalize for the number of parameters, denoted  $n_{\text{pars}}$  as follows:

$$\begin{aligned} \text{AICc} &= -2 LL^* + 2 n_{\text{pars}} + \frac{2 n_{\text{pars}}(n_{\text{pars}} + 1)}{n_{\text{trials}} - n_{\text{pars}} - 1} \\ \text{BIC} &= -2 LL^* + n_{\text{pars}} \log n_{\text{trials}} \end{aligned}$$

While the AICc penalty might be not harsh enough with increasing number of parameters, the BIC penalty might be a bit too harsh in our case, since there are parameter trade-offs.

## 2.4 Results Experiment 1

### 2.4.1 Target localization

The goal of the experiment was to measure the effects of set size, task and time on visual search with heterogenous distractors. To begin, we focus on the localization data in isolation; n-AFC localization with heterogenous distractors has not been studied with an optimal-observer approach.

#### 2.4.1.1 Summary statistics

What limits performance in the localization of an orientation target in perception and memory? We can think of each Gabor stimulus as a point in a 2D space of continuous orientation and discretized locations. We expect that orientation similarity of the target relative to the distractors should influence performance, and also ask whether spatial distance would additionally influence performance.

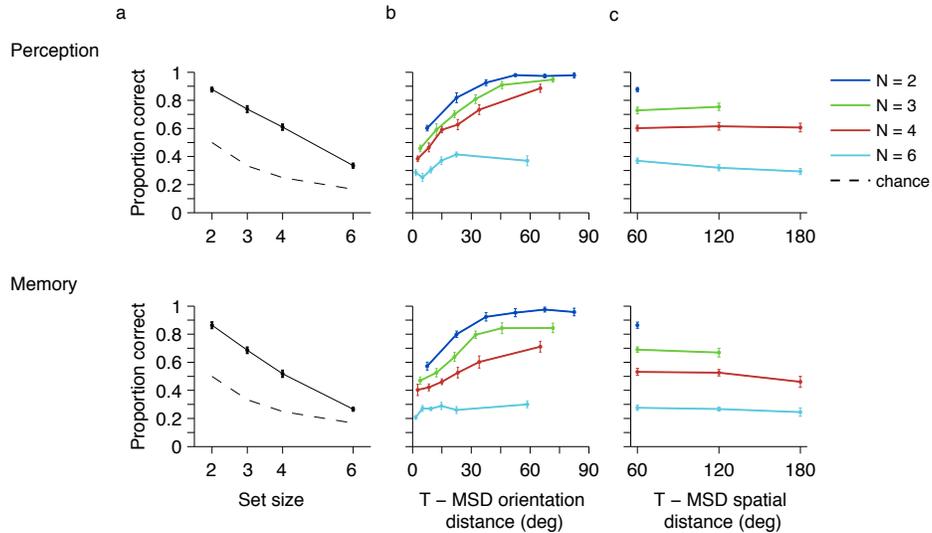
To quantify how confusable trials are with respect to the feature of interest, orientation, we take the circular distance from the target orientation to the orientation of the most similar distractor (MSD). Since the distribution of the (T - MSD) circular distances varies with set size, we placed the bin values separately for each set size (Figure 2.5b), such that equal number of data points would go into each bin. Taking into account that stimuli are uniformly distributed, we analytically calculated the positions of these bins (see Supplementary Section 2.7.1).

As expected, proportion correct drops with set size and with lower distance between

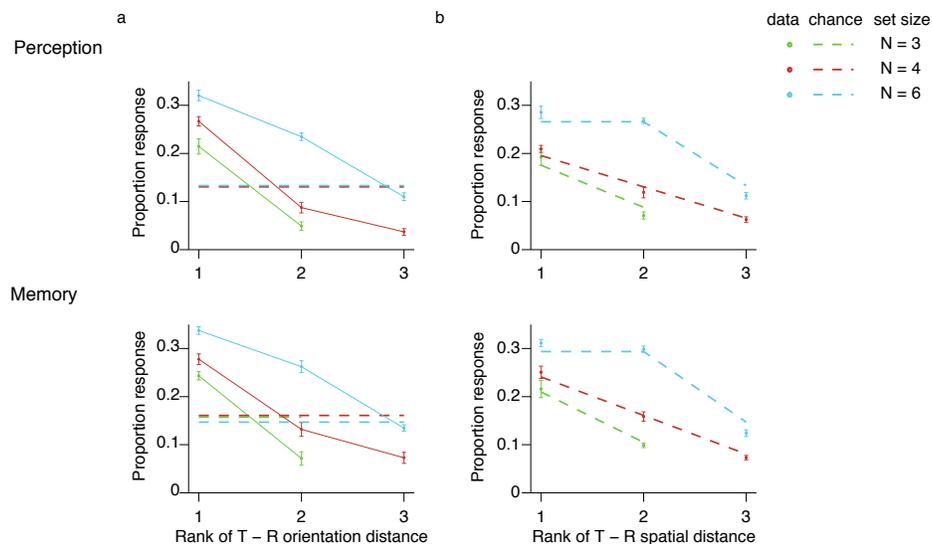
the target and MSD (Figure 2.5 a and b). When looking into the spatial distances of the MSD relative to the target, we do not see that performance is further affected in trials in which the MSD is located closer in space (Figure 2.5c). This might have been expected under the interference model of (Oberauer and Lin, 2017) or if swaps were to also increase with spatial distance in addition to orientation distance.

When participants are responding incorrectly, what seems to drive these response choices? Proportion response decreases with the increase of the rank of the target - response (T-R) orientation distance, but not with the rank of the T-R spatial distance (Figure 2.6). Thus, while responses seem to be biased towards the target with respect to orientation distance, there is no such bias with respect to spatial distance (the proportion of responses show a decreasing trend here as well, but comparably to how the chance line decreases). This pattern reinforces the earlier intuition that if swaps were to occur, they would not be driven by spatial distances.

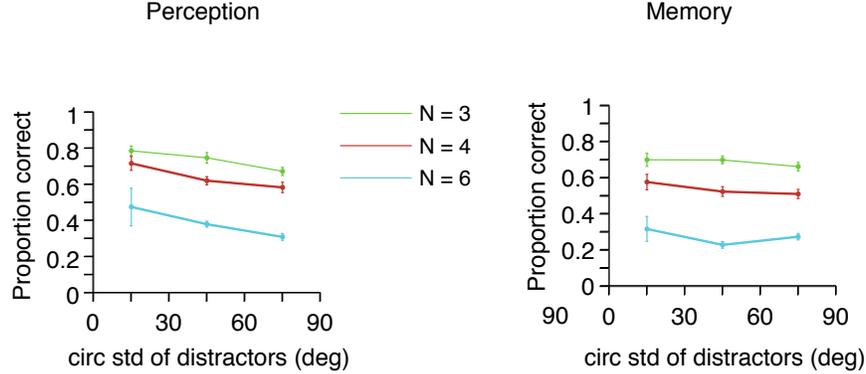
For the last summary statistic, we make use of the heterogeneity of the distractors. Specifically, we calculated on a trial by trial basis the standard deviation of the given distractors. We only show trials with set size greater or equal than 3. With an overall coarse analysis, we do not see an effect of the heterogeneity of the distractors on proportion correct (Figure 2.7). This is in contrast with (Duncan and Humphreys, 1989) that predicted a decrease in proportion correct with increasing distractor heterogeneity. However, it is possible that further splitting of the data according to both T-MSD distance and distractor heterogeneity might yield a different effect, akin to the interaction suggested by (Duncan and Humphreys, 1989).



**Figure 2.5** Localization proportion correct depends on set size and the orientation distance from target to the most similar distractor (MSD). **(a)** Proportion correct decreases as a function of set size. Here and elsewhere, values represent means across participants and error bars the standard error of the means (sem). **(b)** For each set size, proportion correct increases with the orientation distance of the target to the MSD. **(c)** The variation of proportion correct with the spatial distance of the target to the MSD seems to be overlapping with the chance line (dashed line). These metrics show similar trends for **(Top)**: Perception and **(Bottom)**: Memory.



**Figure 2.6** Breakdown of error responses by similarity of response (R) to target (T). **(a)** When people choose the incorrect response, they tend to choose the item that is more similar to the target in orientation distance - data and model fits show decreasing trends while the dashed line representing chance is flat. **(b)** In contrast, the way in which the proportion of responses depends on the rank of the T-R spatial distance seems to be comparable to chance (dashed lines). These metrics show similar trends for **(Top)**: Perception and **(Bottom)**: Memory.



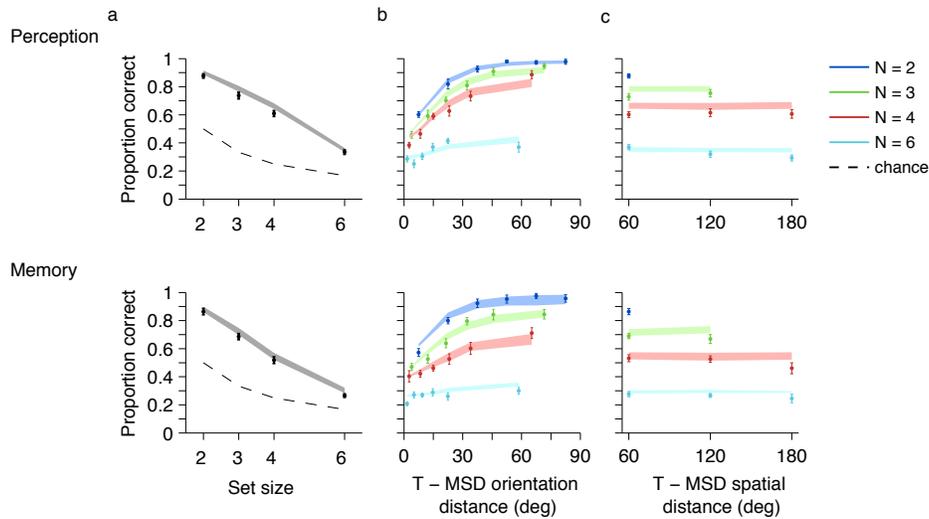
**Figure 2.7** Proportion correct as a function of the circular standard deviation of the distractors. There seems to be no degradation of proportion correct as distractor heterogeneity increases. We see similar trends for **(Left)**: Perception and **(Right)**: Memory.

#### 2.4.1.2 Optimal-observer model fits

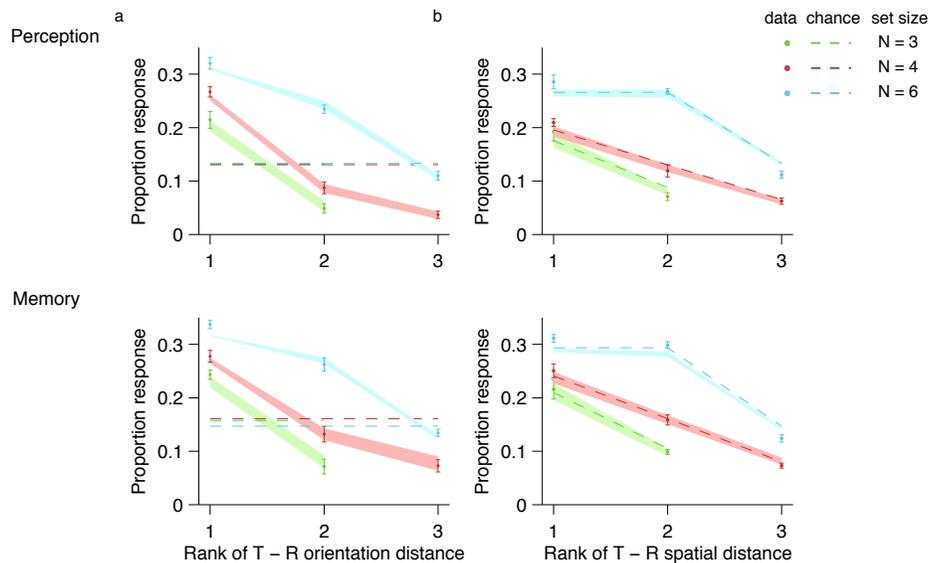
We see that the proportion correct data varies steeply with target-MSD orientation distance, suggesting that an optimal-observer model with a variable precision encoding stage might be able to account for the localization performance data. We first fit the NoLocLapse model described in Section 2.3.1 and see that it visually captures the data well (Figures 2.8 and 2.9).

#### 2.4.1.3 Model comparison

While NoLocLapses (NL) is a qualitatively good fit, we additionally fit the alternative models with lapses LocLapse1 (L1) and LocLapse2 (L2) (shown in Table 2.1) and do quantitative model comparison. We used the corrected Akaike Information Criterion (AICc) and the Bayesian Information criterion (BIC) (described in Section 2.3.2.2) to

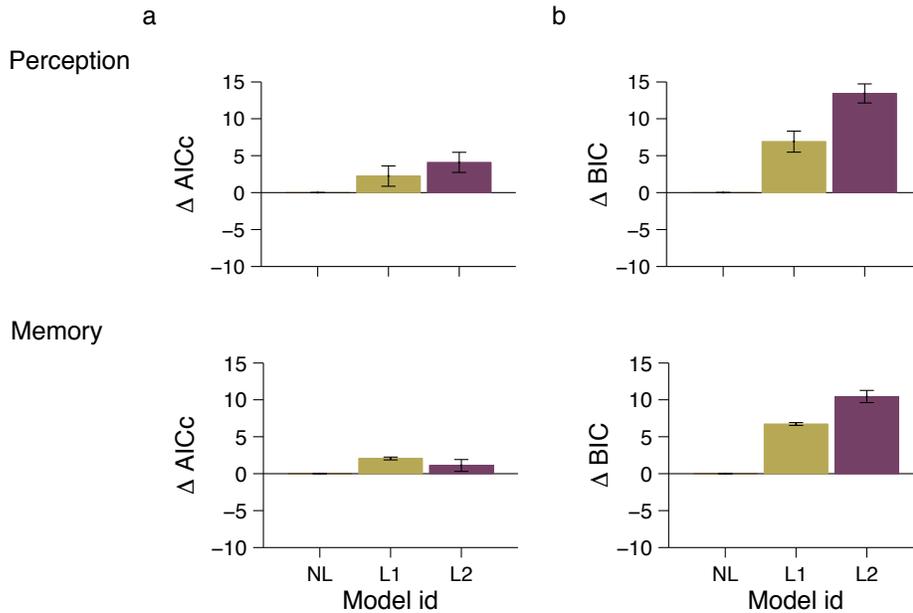


**Figure 2.8** Localization proportion correct, data and model fits. **(a)** Proportion correct with set size. **(b)** Proportion correct with the orientation distance of the target to the most similar distractor (MSD). **(c)** Proportion correct with the spatial distance of the target to the MSD. **(Top):** Perception and **(Bottom):** Memory



**Figure 2.9** Breakdown of error responses by similarity of response (R) to target (T), with model fits. **(a)** When people choose the incorrect response, they tend to choose the item that is more similar to the target in orientation distance - data and model fits show decreasing trends while the dashed line representing chance is flat. **(b)** In contrast, the way in which the proportion of responses depends on the rank of the T-R spatial distance seems to be comparable to chance (dashed lines). These metrics show similar trends for **(Top):** Perception and **(Bottom):** Memory.

compare the NL fits with the fits of L1 and L2. We found that indeed the NL model fitted better than both L1 (in median by about 2 according to AICc and by 6 according to BIC, for both perception and memory) and L2 (in median by 4 for perception and 3 for memory according to AICc and by 13 and respectively 12 according to BIC) (Figure 2.10).

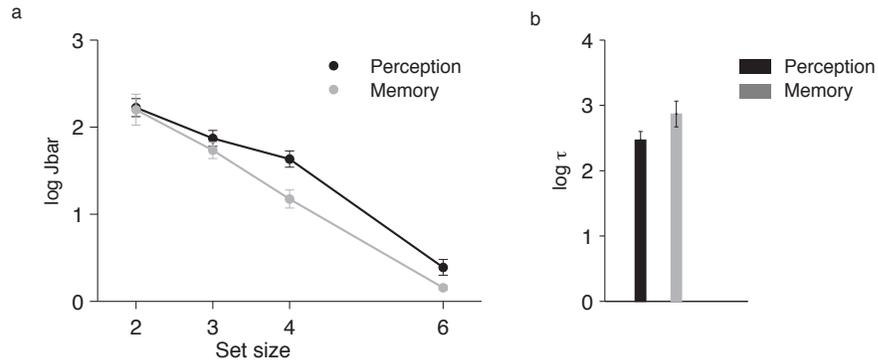


**Figure 2.10 Model comparison: NoLocLapse captures the data best according to both metrics, for both Perception and Memory. (a) Model NL has the lowest AICc (b) Same for BIC.**

#### 2.4.1.4 Mean precision parameters across set sizes and perception vs memory

Mean precision,  $\bar{J}$  seems to decrease with set size and has higher values for perception than memory (Figure 2.11). A 2-way repeated measures ANOVA on  $\log \bar{J}$  showed a significant effect of set size ( $F(3, 30) = 203, p < 0.0001$ ) and a significant effect of time ( $F(1, 10) = 9, p = 0.01$ ), with a borderline significant interaction between them ( $F(3, 30) = 3, p = 0.05$ ). The  $\tau$  parameter, measuring the spread of the gamma distribution of precisions, is different for perception and memory (Wilcoxon sign-rank  $p = 0.01$ ), suggesting

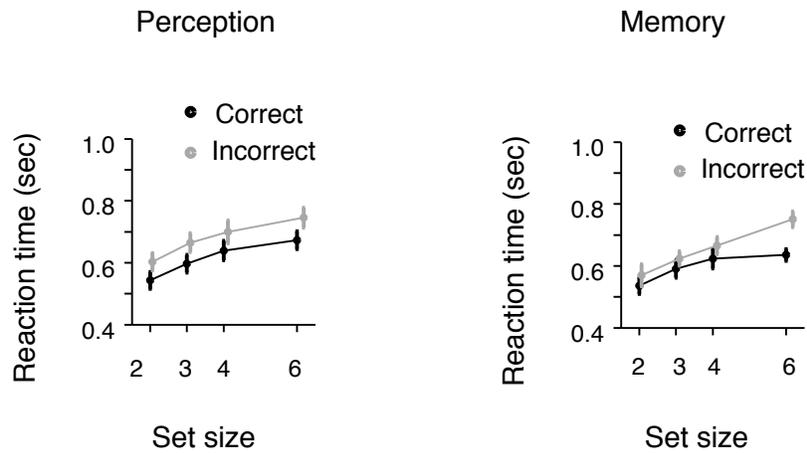
wider distributions in the memory condition. In the VP encoding model, trade-offs between  $\bar{J}$  and  $\tau$  are a possible concern, so it is good to see that the higher  $\bar{J}$  in perception is not accompanied by a higher  $\tau$ ; in that case we would not have been able to conclude so readily that precision is higher in perception than memory.



**Figure 2.11 Localization parameters for Perception and Memory. (a)** Mean precision  $\bar{J}$  with set size. **(b)**  $\tau$  parameter.

#### 2.4.1.5 Reaction times

We also show the reaction times - reaction times increase with set size and tend to be higher for incorrect (Figure 2.12), sign that speed - accuracy trade-offs are not a concern in our data.



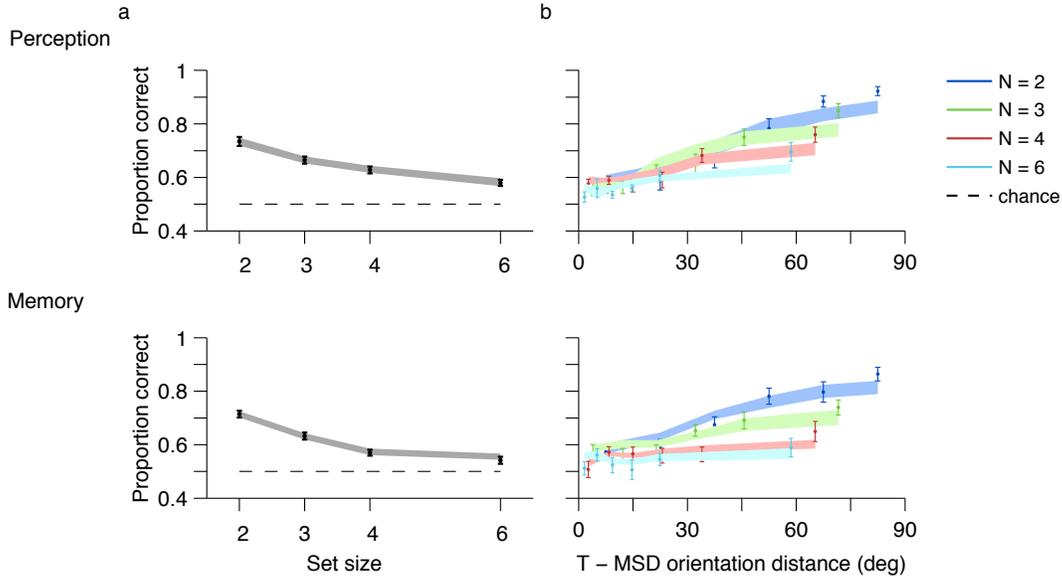
**Figure 2.12 Localization reaction times.** Dots represents means across observers of the RT medians across the selected trials, and error bars sem across observers. RT with set size, separately for correct and incorrect.

## 2.4.2 Target detection

We now also include the second task in our dataset, detection. We first look at the optimal-observer model fits to detection in isolation, recapitulating the study of (Mazyar et al., 2012).

### 2.4.2.1 Optimal-observer model fits

As expected, we were able to capture their data well with the optimal-observer model. These fits, while satisfactory, could be better.



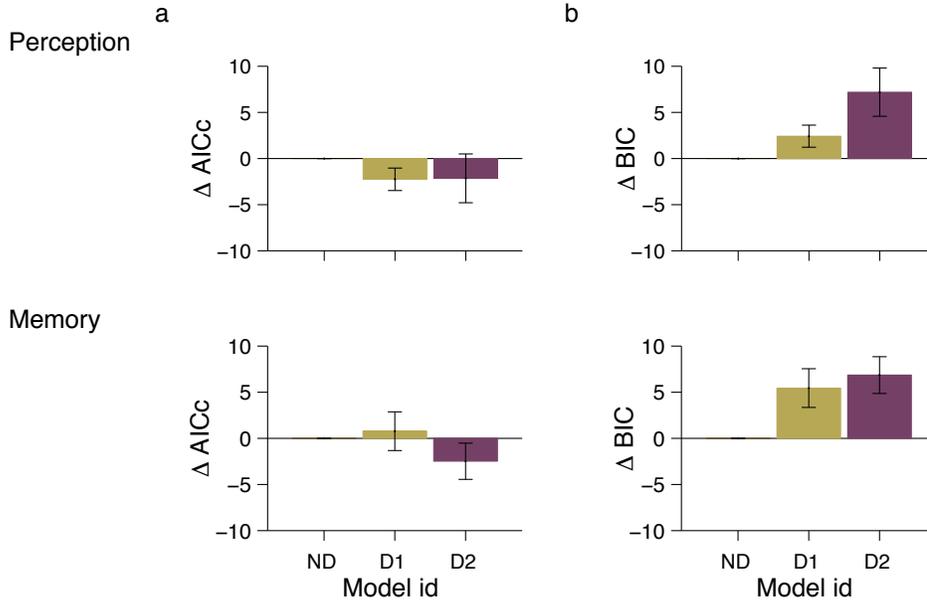
**Figure 2.13** Detection proportion correct, data and model fits. **(a)** Proportion correct with set size. **(b)** Proportion correct with the orientation distance of the target to the most similar distractor (MSD).

#### 2.4.2.2 Model comparison

Adding a generic lapse (D1) seems to improve the model fits for both memory and perception according to AICc, but not according to BIC. Thus, we do not find consistent evidence that adding lapses to ND substantially improves the model fit.

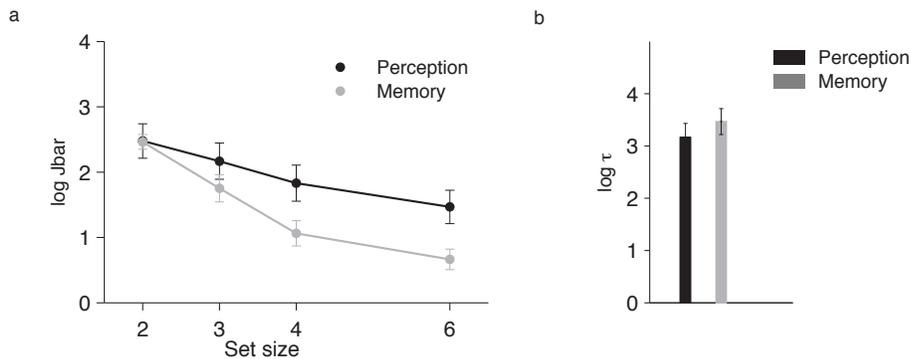
#### 2.4.2.3 Mean precision parameters across set sizes and perception vs memory

Mean precision,  $\bar{J}$  decreases with set size and tends to have higher values for perception than memory, as expected. A 2-way repeated measures ANOVA on  $\log \bar{J}$  showed a significant effect of set size ( $F(3, 30) = 13.6, p < 0.0001$ ) and a significant effect of time ( $F(1, 10) = 11.1, p = 0.0075$ ), with a significant interaction between them ( $F(3, 30) = 4.6, p = 0.009$ ). The  $\tau$  parameter, measuring the spread of the gamma distribution of



**Figure 2.14 Model comparison:** There is no consistent evidence that adding parameters to NoDetLapse improves the model fit. (a) AICc (b) BIC.

precisions, is not significantly different for perception and memory (Wilcoxon sign-rank  $p = 0.41$ ), Note that the reliability of these parameter estimates should be taken more cautiously since the model fits are not quite as good as the localization data model fits.



**Figure 2.15 Detection parameters for Perception and Memory.** (a) Mean precision with set size. (b)  $\tau$  parameter.

### 2.4.3 Target localization and target detection

#### 2.4.3.1 Joint localization and detection model fits

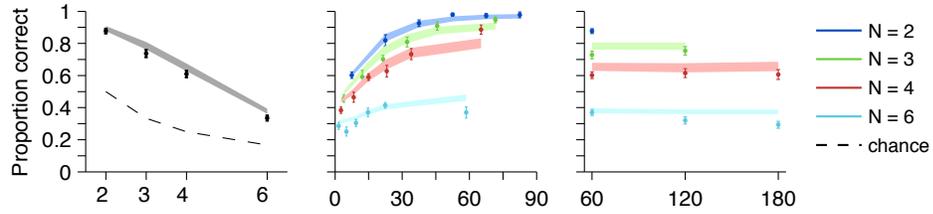
Since the same observers performed both the detection and the localization tasks, we can jointly fit the data from both tasks per observer separately per perception and memory conditions. The generative models are highly similar in the encoding stages and only differ in the decision rules. As we see in Figure 2.16, the joint model fits are pretty good.

#### 2.4.3.2 Joint mean precision parameters

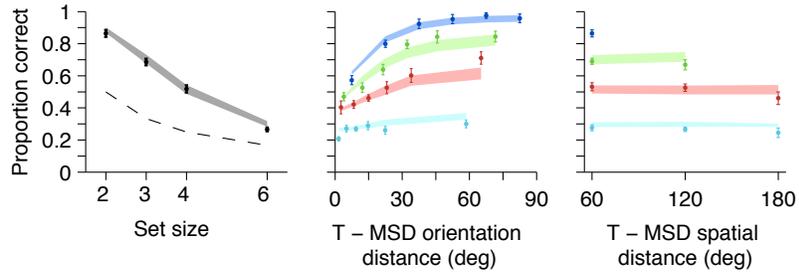
As in the case of Localization data alone, mean precision,  $\bar{J}$  decreases with set size and has higher values for perception than memory, as expected. The  $\tau$  parameters in perception are not statistically different than the ones in memory ( $p = 0.36$ ). Since these parameter estimates and patterns might change after future attempts to better fit the detection data, we do not delve further into these parameter estimates at this point.

### Localization

Perception

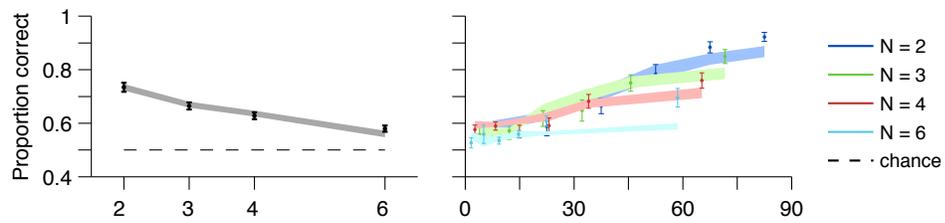


Memory



### Detection

Perception



Memory

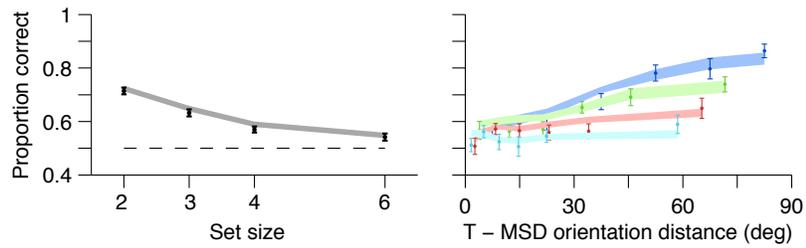
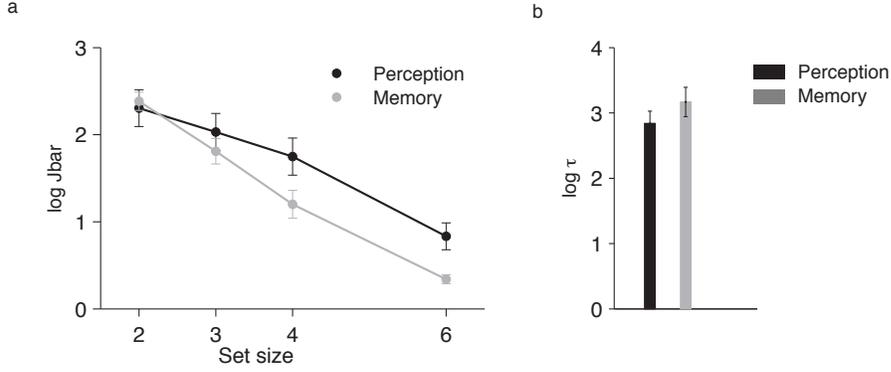


Figure 2.16 Localization and Detection - joint model fits capture the data pretty well.



**Figure 2.17** Joint localization and detection parameters for Perception and Memory. **(a)** Mean precision with set size. **(b)**  $\tau$  parameter.

## 2.5 Results of Experiment 2: effect of stimulus spacing

The structure of this section recapitulates the structure of Section 2.4.

### 2.5.1 Target localization

Before delving into the model-based results of Experiment 2, we compared the data across Experiments 1 and 2 with a mixed-design ANOVA, with the between-group factor Spacing (Experiment 1: 60 deg and Experiment 2: 30 deg). This will give us an initial sense whether bringing the stimuli closer together, though still outside of the Bouma critical distance for crowding, would lead to degradation in performance. A 3-way mixed-design ANOVA for the Localization proportion correct data with the between-group factor Spacing and the within-group factors Timing - Perception and Memory - and Set size - 2,3,4 and 6 yielded significant effects of Timing ( $F(1, 16) = 66, p < 0.0001, \eta_p^2 = 0.81$ ) and Set Size ( $F(3, 48) = 1675, p < 0.0001, \eta_p^2 = 0.99$ ), but not of Spacing ( $F(1, 16) = 0.4, p = 0.55, \eta^2 = 0.02$ ). We saw significant interactions of Spacing  $\times$  Set Size ( $F(3, 48) = 4.6, p =$

0.016,  $\eta^2 = 0.23$ ) and Timing  $\times$  Set Size ( $F(3, 48) = 7.8, p = 0.001, \eta^2 = 0.30$ ). Thus, bringing the stimuli closer together in Experiment 2 does not seem to lead to significant changes in proportion correct.

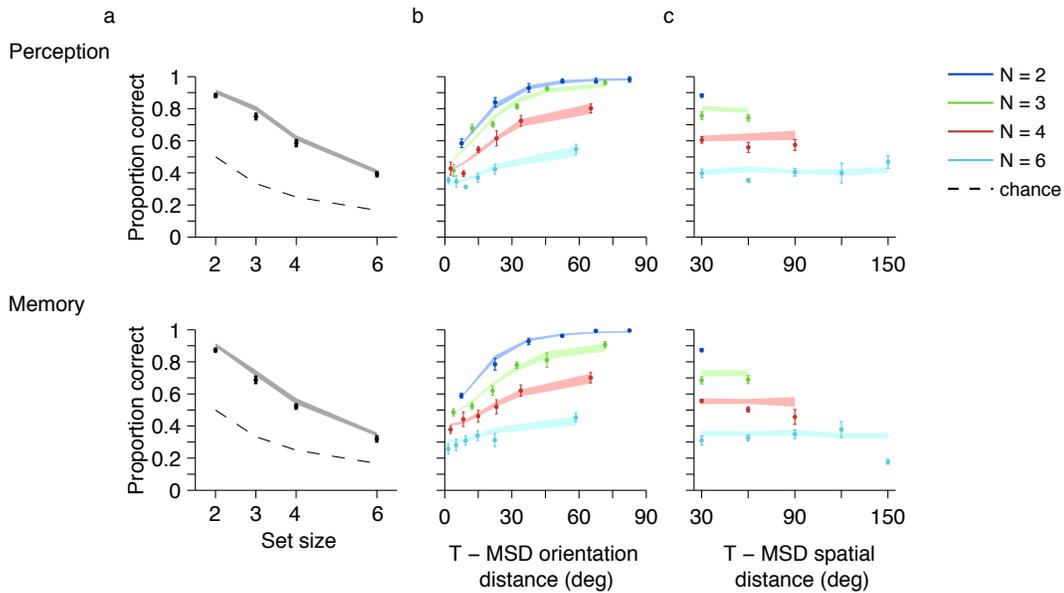
As in Experiment 1, we see that the proportion correct data varies steeply with the orientation distance of the target to the most similar distractor (MSD) and that an optimal-observer model with a variable precision encoding stage accounts well for the localization performance data.

Again, as in Experiment 1, with an overall coarse analysis, we do not see an effect of the standard deviation of the distractors on proportion correct (Figure 2.20). Model predictions seem to follow the same pattern. This is in contrast with (Duncan and Humphreys, 1989) that predicted a decrease in proportion correct with increasing distractor heterogeneity. Again, we cannot rule out that further splitting of the data according to both T-MSD distance and distractor heterogeneity might yield a different effect.

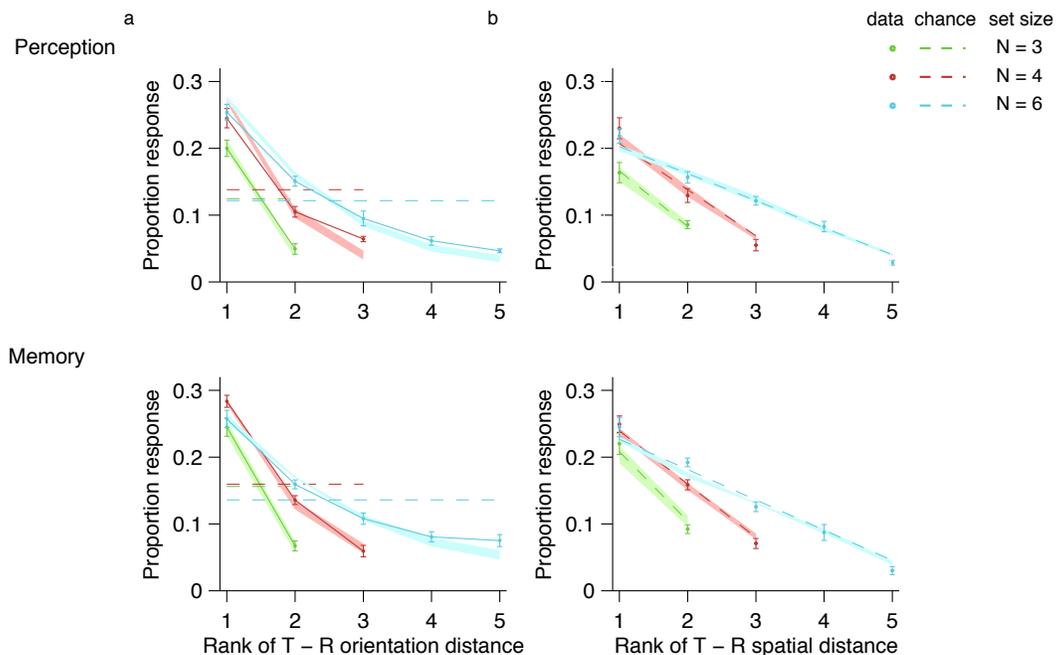
As in Experiment 1, we found that indeed the NL model fitted better than both L1 and L2, suggesting that there is no benefit to adding lapse parameters.

As in experiment 1, mean precision,  $\bar{J}$  decreases with set size and has higher values for perception than memory. A 2-way repeated measures ANOVA on  $\log \bar{J}$  showed a significant effect of set size ( $F(3, 18) = 19, p < 0.0001$ ) and a significant effect of time ( $F(1, 6) = 22, p = 0.003$ ), and a significant interaction set size  $\times$  time ( $F(3, 18) = 17, p < 0.001$ ). The  $\tau$  parameters, measuring the spread of the gamma distribution of precisions do not seem different between the perception and memory conditions (Wilcoxon sign-rank  $p = 0.9$ ).

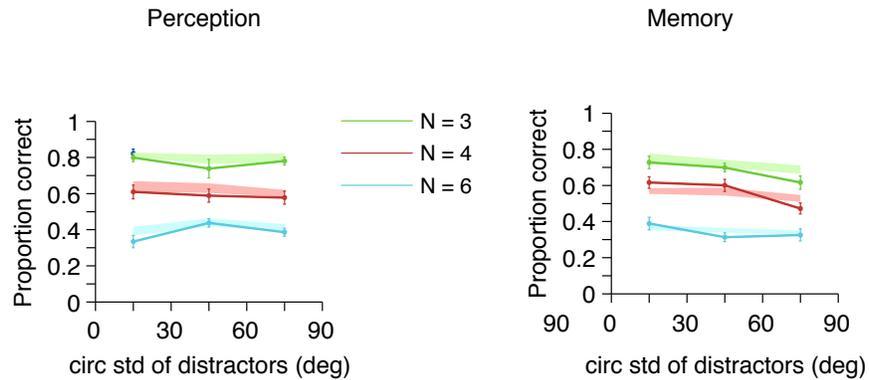
While the precision parameters here seem lower than the ones in Experiment 1, so are the



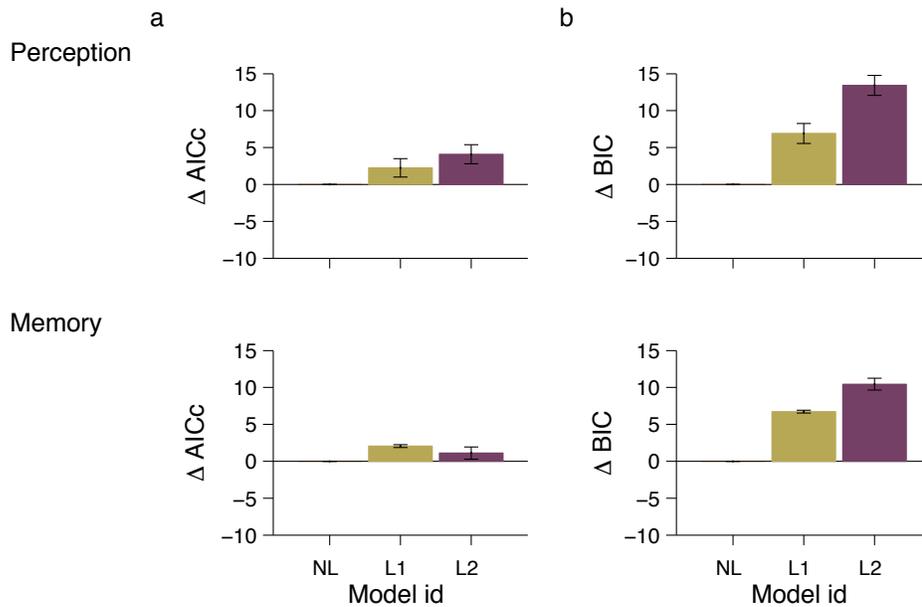
**Figure 2.18 Experiment 2. Localization proportion correct, data and model fits. (a)** Proportion correct with set size. **(b)** Proportion correct with the orientation distance of the target to the most similar distractor (MSD). **(c)** Proportion correct with the spatial distance of the target to the MSD. **(Top):** Perception and **(Bottom):** Memory



**Figure 2.19 Breakdown of error responses by similarity of response (R) to target (T), with model fits. (a)** When people choose the incorrect response, they tend to choose the item that is more similar to the target in orientation distance - data and model fits show decreasing trends while the dashed line representing chance is flat. **(b)** In contrast, the way in which the proportion of responses depends on the rank of the T-R spatial distance seems to be comparable to chance (dashed lines). These metrics show similar trends for **(Top):** Perception and **(Bottom):** Memory.

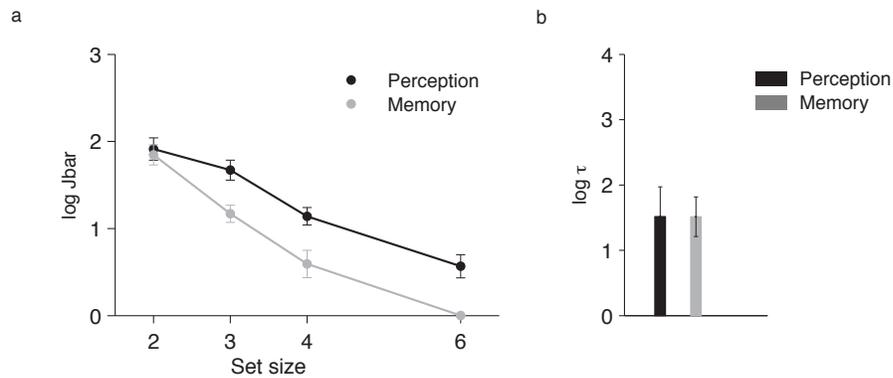


**Figure 2.20** Proportion correct as a function of the circular standard deviation of the distractors, overlaid with model predictions. There seems to be no degradation of proportion correct as distractor heterogeneity increases and the model seems to capture this. We see similar trends for **(Left)**: Perception and **(Right)**: Memory.



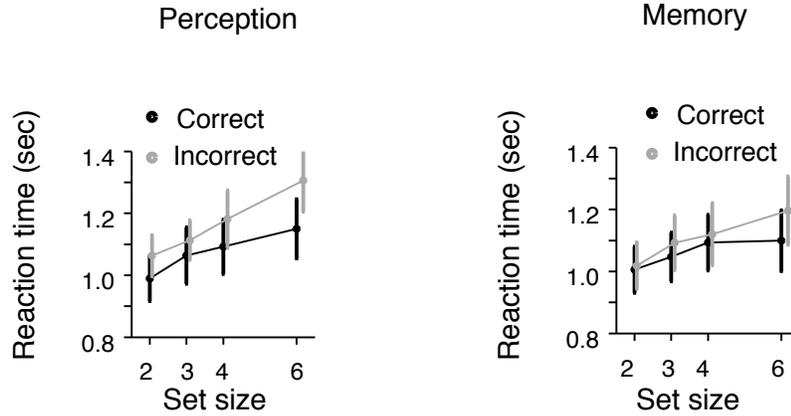
**Figure 2.21** Model comparison: NoLocLapse captures the data best according to both metrics, for both Perception and Memory. **(a)** Model NL has the lowest AICc **(b)** Same for BIC.

$\tau$  parameters; thus, the gamma distributions described by these parameters could be overlapping substantially. Because of this trade-off, we cannot conclude that precision is lower in Experiment 2 relative to Experiment 1.



**Figure 2.22 Localization parameters for Perception and Memory. (a)** Mean precision with set size. **(b)**  $\tau$  parameter.

The reaction time patterns recapitulate the patterns in Experiment 1, though note the values on the y axes have substantially higher values (0.8 to 1.4 sec here vs 0.4 to 1.0 sec in Experiment 1). Thus, even if participants achieve comparable proportion correct and possible comparable precision parameters, they seem to do so at the expense of an increase in reaction times. A 3-way mixed-design ANOVA for the reaction time Localization data with the between-group factor Spacing and the within-group factors Timing - Perception and Memory - and Set size - 2,3,4 and 6 yielded significant effects of Timing ( $F(1, 16) = 5.8, p = 0.028, \eta_p^2 = 0.27$ ) and Set Size ( $F(3, 48) = 23.5, p < 0.0001, \eta_p^2 = 0.83$ ), and of Spacing ( $F(1, 16) = 33.3, p < 0.0001, \eta^2 = 0.67$ ).



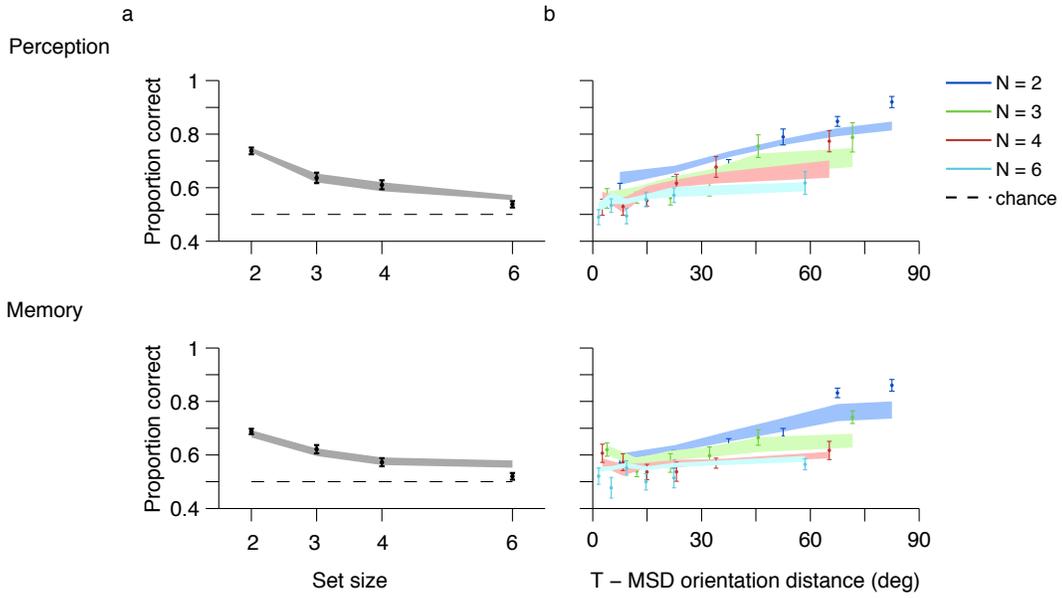
**Figure 2.23** Localization reaction times. RT with set size, separately for correct and incorrect.

### 2.5.2 Target detection

Just like for Localization, before delving into the model-based analysis, we carried out a 3-way mixed-design ANOVA with between-group factor Spacing and the within-group factors Timing - Perception and Memory - and Set size - 2,3,4 and 6. We saw significant effects of Timing ( $F(1, 16) = 27.6, p = 0.0001, \eta_p^2 = 0.6$ ) and Set Size ( $F(3, 48) = 99.2, p < 0.0001, \eta_p^2 = 0.88$ ), but not of Spacing ( $F(1, 16) = 1.18, p = 0.3, \eta_p^2 = 0.08$ ). None of the interactions were significant ( $p > 0.25$ ).

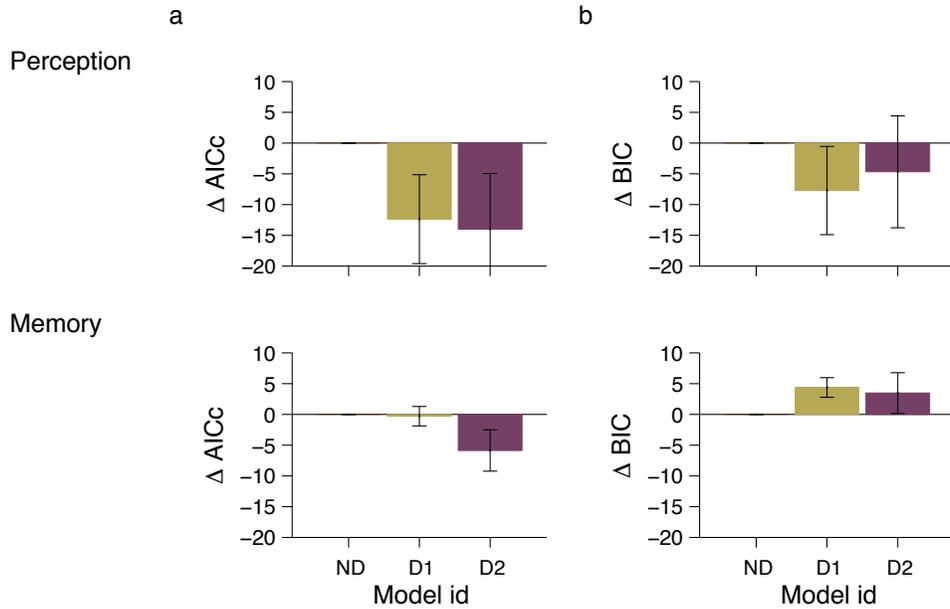
As in Experiment 1, the same observers also performed the target detection task. Based on Experiment 1 and (Mazyar et al., 2012), we expected a better match between the data and the model.

Adding a generic lapse (D1) seems to improve the model fits for both memory and perception according to AICc, but not according to BIC. Thus, though less convincingly than in Experiment 1, we do not find consistent evidence that adding lapses to ND sub-



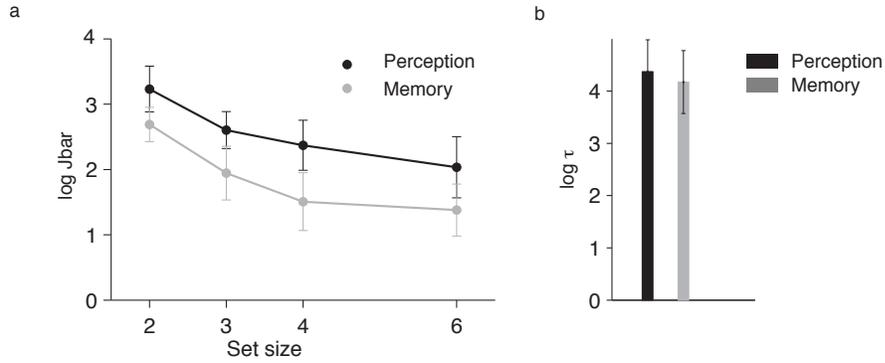
**Figure 2.24** Detection proportion correct, data and model fits. **(a)** Proportion correct with set size. **(b)** Proportion correct with the orientation distance of the target to the most similar distractor (MSD).

stantially improves the model fit.



**Figure 2.25** Detection: model comparison. **(a)** AICc **(b)** BIC.

As in Experiment 1, mean precision,  $\bar{J}$  decreases with set size and has higher values for perception than memory, as expected.



**Figure 2.26** Detection parameters for Perception and Memory. (a) Mean precision with set size. (b)  $\tau$  parameter.

### 2.5.3 Target localization and target detection

As we see in Figure 2.27, the joint model fits are good. The ability of this joint model to capture the data in detection is not great, as it was the case when we fitted the detection data alone.

As in the case of Experiment 1, mean precision,  $\bar{J}$  decreases with set size and has higher values for perception than memory. The  $\tau$  parameters seem comparable. As we said for experiment 1, since these parameter estimates and patterns might change after future attempts to better fit the detection data, we do not delve further into these estimates at this point.

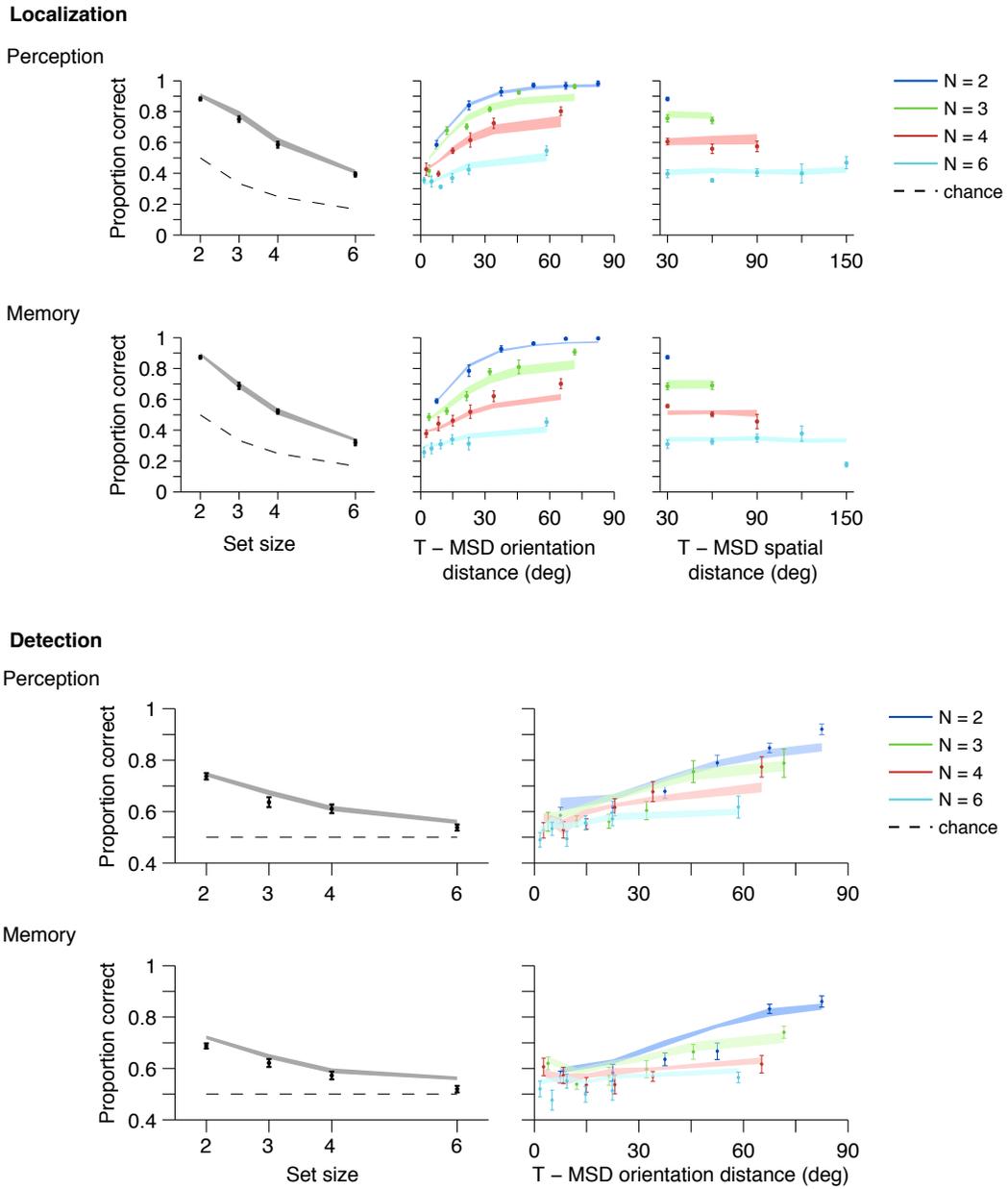
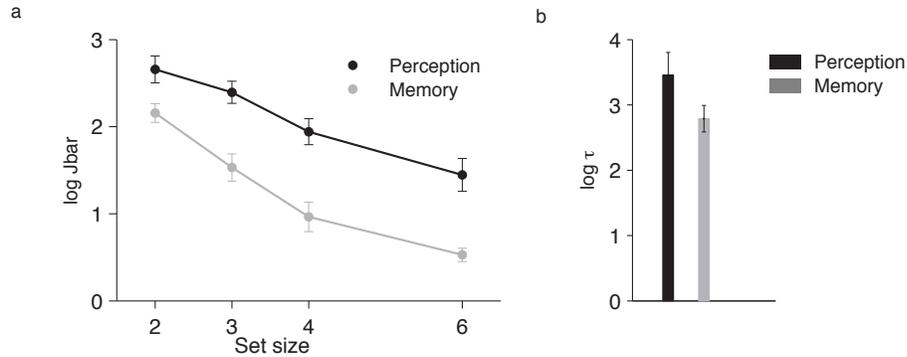


Figure 2.27 Localization and Detection - joint model fits capture the data well.



**Figure 2.28** Joint localization and detection parameters for Perception and Memory. **(a)** Mean precision with set size. **(b)**  $\tau$  parameter.

## 2.6 Discussion

Here, we extended the study of visual search to heterogenous distractors; specifically, we characterize the effects on performance of set size, task - detection and localization, time (perception and memory) and space. We used an experimental design that maximized the chance that we were engaging divided attention: brief stimulus presentation times to prevent observers from making sequential eye movements to different items and placement of stimuli in the search array at the same eccentricity with wide spacing attempting to prevent crowding phenomena (Palmer, 1995, 2014; Palmer et al., 1993).

As expected, we found that percent correct decreased with the set size of the search array for both detection and localization. Proportion correct also degraded with increasing target-most similar distractor (MSD) orientation distance, but not with target-MSD spatial distance. We found that proportion correct did not seem to depend on distractor heterogeneity, as quantified with the standard deviation of the distractors calculated on a trial by trial basis. An optimal observer model with a variable precision encoding of orientation stage and an optimal decision rule was able to capture behavior in a more natu-

realistic task, target localization, without the need to build in loss of location information. We found that precision decreased with set size for both localization and detection. This is not an obvious result, as some pattern of decrease in performance with set size could have been accomplished with mean precisions that are constant with set size (Mazyar et al., 2012); the decrease in performance would have been due to the fact that the more distractors there are, the higher the chance of one of them to be confused with the target, as in the signal detection theory account (Palmer, 1995). As expected, mean precision was generally higher in the perception condition relative to the memory condition. Joint fits for localization and detection data were good, suggesting that observers might be using the same encoding processes across the two tasks, as well as possibly the respective decision rules derived from the ideal-observer models. There is some doubt mainly since the detection fits could have been better. We found the same pattern of results with visual search arrays with reduced stimulus spacing; observers were able to achieve comparable precision parameters, albeit with increased reaction times.

We were able to capture localization data with an optimal-observer model with variable precision encoding and Bayesian decision rule; the observers' performance depended on the T-MSD orientation distance, but not on T-MSD spatial distance, and our model captured that, without the need for additional lapse parameters. This is interesting to discuss in light of the results of (Bays, 2016; Bays et al., 2009) in which, together with feature noise, an additional mechanism had to be included to account well for observers' distribution of errors in delayed estimation tasks. This additional mechanism produces errors called "swaps" or non-target reports or misbinding errors, which entail reporting the identity of another item in the array, as if observers swapped the identity of the target item with the identity of an item at another location. If swaps would have happened in our localization task, they would have been incorporated in the lapse param-

ters, but there was no need for lapse parameters. Several reasons could be at the root of this discrepancy. First, task instructions likely played a role: participants might have remembered locations in our task since they were explicitly asked to report them, while in delayed estimation they had to report feature values exclusively. Additionally, the dorsal stream engaged in a perception for action task such as localization could have higher spatial resolution (Geniva et al., 2003). Lastly, their item arrays were placed at arbitrarily close locations, while we had regularity since we always placed the items 60 dva apart on a concentric circle in Experiment 1 and 30 dva apart in Experiment 2. Overall, it is likely that we found that observers can encode and remember location information well if it is task relevant as it is in localization; therefore, we provide an upper bound for the observers' ability to encode locations as well. To more directly compare to the (Bays, 2016; Bays et al., 2009), future studies could take inspiration from (Shin and Ma, 2016), who harnessed the power of Amazon Mechanical Turk to give participants on the last trial a surprise probe about the task irrelevant feature (orientation or color). A similar design could address more definitively whether location would be remembered even when it is not explicitly relevant to the task.

We fitted the optimal-observer model with shared parameters to localization and detection data jointly. Our work connects to some extent to joint investigations of detection and localization from visual search with homogenous distractors. (Cameron et al., 2004) showed that signal detection theory can explain performance as a function of set size for identification, detection and localization. (Geniva et al., 2003) asked whether detection and localization are underlined by either a shared process or two distinct processes, potentially accomplished by the ventral and respectively dorsal streams. In our data, while the joint detection and localization fits could be better, the fact that they capture the data as well as they do seems to add further weight to the assumption of shared variable precision

in the early visual encoding processes in target detection and target localization. It is possible, however, that the decision processes in localization vs detection could be different. The detection fits could be improved, especially in Experiment 2; perhaps people are using a suboptimal decision rule, such as a max rule applied to a threshold. Alternatively, it is possible that decision noise might play a role in our dataset. There was no advantage of adding a decision noise parameter to the (Mazyar et al., 2012) data, as showed by (Shen and Ma, 2017), but we could get a different result in our dataset.

We were able to capture the localization data very well with an optimal-observer model with variable precision encoding and Bayesian decision rule. But perhaps a precision that has a set-size specific value but does not further vary across trials and items would be sufficient to capture the data comparably well with one less parameter. It would be useful to also test an encoding scheme with a precision distribution that has the same mean for every set size, but is allowed to vary across trials and items as captured by a  $\tau$  parameter. These different encoding schemes have been investigated in previous visual search work and found to be worse than variable precision (Mazyar et al., 2012, 2013; Shen and Ma, 2017), but we should also extend this thorough testing to localization. Or perhaps suboptimal decision rules would also capture the data well. Thus, immediate future work entails a factorial model comparison scheme including other ways to parametrize encoding precision and also suboptimal decision rules. This could reveal the source of the deviations from the optimal-observer model we examined in the detection data, as well as test whether this particular optimal-observer model is still the best account of the localization data, upon competitive model comparison.

Once we will have increased confidence that we found the best fitting models through competitive model comparison, we could further compare precision parameters (or distri-

butions) across far vs near stimuli spacing. In the case of localization, even if the fits were good, it was not immediate to do so due to parameter trade-offs. The mean precision values were larger for far stimuli, but so were the  $\tau$  parameters. It would be interesting to see if there are different patterns in perception vs memory for the possible decay of precision with increased stimulus spacing. Since even our near stimuli are outside Bouma’s limit for crowding, our results would complement the study of (Tamber-Rosenau et al., 2015), which focused on the crowding regime and showed comparable spatial resolution of representations in perception and working memory.

## 2.7 Supplementary

### 2.7.1 Calculation of quantile bins for the psychometric curves

We plot the psychometric curves metrics as a function of the minimum orientation difference between target and distractors, which represents a measure of the differentiability between the target and the distractors. We denote the minimum orientation difference distribution  $p(\theta)$ . Its cumulative density function (cdf) is  $P(\min a_i < \theta)$ . Quantile binning of this distribution is important to ensure that there will be enough data points in each bin and thus the psychometric curve is reliable. The quantile function is the inverse of the cdf, which we can analytically compute.

The cdf will vary with condition and set size. The search arrays have set sizes  $N = 2, 3, 4, 6$ , and with conditions target present and target absent, the absolute orientation differences  $a_i, i = 1 - N$  could have length anywhere between  $N = 1 - 6$ . Both foveally presented targets and peripherally presented search array stimuli are independently drawn from uni-

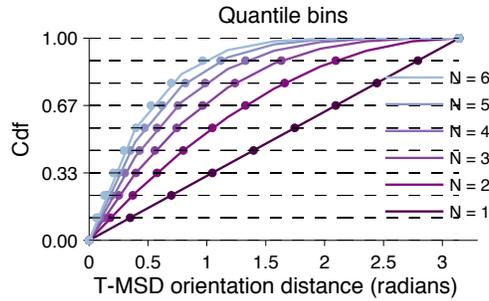
form distributions on  $(-\pi, \pi)$ , so  $a_i \in (0, \pi)$ .

We will decompose the cdf  $P(\min a_i < \theta)$  as below.  $\min a_i < \theta$  is contradicted when  $\forall a_i > \theta$ , for  $i = 1 - N$ , which are independent and thus we write as a product over  $i$ . Since the search arrays have set sizes  $N = 2, 3, 4, 6$  with both target present and target absent conditions, we could have anywhere between  $N = 1 - 6$  angle differences  $a_i$ .

$$\begin{aligned}
 P(\min a_i < \theta) &= 1 - P(a_i > \theta \text{ for all } i) = 1 - \prod_{i=1}^N P(a_i > \theta) \\
 &= 1 - \prod_{i=1}^N \left(1 - \int_0^\theta p(a_i) da_i\right) = 1 - \prod_{i=1}^N \left(1 - \int_0^\theta \frac{1}{\pi} da_i\right) \\
 &= 1 - \left(1 - \frac{\theta}{\pi}\right)^N
 \end{aligned} \tag{2.4}$$

We show 9 quantiles  $p = [0, 1/9, 2/9 \dots 1]$  in Figure 2.29. More generally, for quantiles  $p \in (0, 1)$ , we solve for the quantile orientation values  $\theta_p$  that satisfy  $P(\min a_i < \theta_p) = p$  to get:

$$\theta_p = \pi \left(1 - (1 - p)^{1/N}\right)$$



**Figure 2.29** Target to most similar distractor (MSD) orientation distance distributions and quantile placement of bins.

# Chapter 3

## Perception, selective attention and task-switching in neurotypicals and ADHD

*People don't change that fast. Labels do.*

Allen J. Frances, 2011

---

### 3.1 Introduction

In Chapter 2, we studied the effects of set size, task, time and space on precision in a visual search experiment in neurotypical participants designed to increase the chance of engaging divided attention. In this chapter, we make use of a selective attention task, with spatial as well as feature dimension switches, and examine both neurotypical and ADHD participants. Beyond having different diagnosis labels, how will the behavior of ADHD participants differ from neurotypicals? Uncovering the attentional deficit in ADHD has

been elusive, at least with classic spatial selective attention paradigms (Huang-Pollock and Nigg, 2003; Roberts et al., 2017). Here we test participants with a more difficult task, that also taxes cognitive control by requiring participants to keep track and switch between stimulus-response rules.

In ADHD, the behavioral deficits captured by self-reports and collateral reports have been attributed to differences in attention, executive function, and lower-level processes, including perceptual function. In the realm of visual attention, differences in accuracy or reaction time have been found in some visual search tasks but not in others (for a review, see (Mullane and Klein, 2008)). No consistent deficits have been found when probing selective attention with visuo-spatial orienting tasks (Cubillo et al., 2010; Huang-Pollock and Nigg, 2003; Roberts et al., 2017; Rubia et al., 2010). ADHD patients tend to have worse executive function than Controls (Boonstra et al., 2005; Castellanos and Tannock, 2002; Kofler et al., 2013; Willcutt et al., 2005), predominantly in response execution and inhibition (Barkley, 1997; Booth et al., 2005; Casey et al., 1997), but also in working memory and switching between stimulus-response rules (Cepeda et al., 2000; Hallelund et al., 2012; Homack, 2004; King et al., 2007).

While some researchers believe executive function impairments to be primary in ADHD, others acknowledge that they are neither necessary nor sufficient to cause the disorder (Boonstra et al., 2005; Willcutt et al., 2005). More specifically, yet others suggest that ADHD impairments are a combination of deficits in high-level and "low-level processes" (Castellanos et al., 2008; Gonen-Yaacovi et al., 2016; Killeen et al., 2013; Rommelse et al., 2007; Sergeant et al., 2002; Sonuga-Barke and Castellanos, 2007). These low-level processes entail arousal (Sergeant, 2005), relatedly, accumulation of evidence (Karalunas and Huang-Pollock, 2013), timing (Nigg and Casey, 2005), or reward sensitivity

(see (Ma et al., 2016; Sonuga-Barke, 2003) for reviews). It should be kept in mind that ADHD might be a heterogenous disorder (Fair et al., 2012; Nigg et al., 2005) and different causes might apply to different deficits.

Here, we examine low-level processes related to perceptual encoding. Behavioral studies that examined the quality of perceptual encoding in ADHD in the absence of attentional or executive involvement have found small and inconsistent differences (see (Fuermaier et al., 2017) for a review). On the other hand, other investigations have found evidence for self-reported impairments in perceptual function in ADHD participants (Bijlenga et al., 2017; Micoulaud-Franchi et al., 2015), or in the general population with ADHD traits (Panagiotidi et al., 2018), as well as deficits in color processing and self-reported visual function in ADHD (Kim et al., 2014c). These findings are not necessarily contradictory, as perceptual deficits might emerge when attention or executive function is simultaneously taxed.

Therefore, we believe it is important to use a task that taxes both perceptual function and either attention and/or executive function, but that allows for a dissociation of the respective processes. This dissociation is difficult, as has been described in the study of autism (Robertson and Baron-Cohen, 2017). In ADHD, there have been a few attempts to dissociate perceptual function from attention within a single task (Kim et al., 2014b; McAvinue et al., 2012; Stevens et al., 2012). For example, (Stevens et al., 2012) compared letter displays with or without distractors and found that ADHD participants had lower performance only when distractors were present. However, spatial covert attention was similar across ADHD and controls, leading the authors to suggest that perceptual interference or crowding is increased ADHD.

It is still unknown whether perceptual function is impaired when executive function

is simultaneously taxed. A study by (Friedman-Hill et al., 2010) used a face discrimination task where they probed perceptual noise by manipulating distractor saliency and probed top-down executive control by parametrically manipulating discrimination difficulty. In difficult discriminations, the reaction time difference between high-saliency and low-saliency distractors was comparable in children with ADHD to that in healthy children and adults; however, in easy discriminations, children with ADHD were slower to respond when presented with low-saliency distractors. These results suggest similar perceptual interference due to distractor saliency in ADHD and Controls, but a higher threshold in ADHD for activating executive control of attention. A problem with (Friedman-Hill et al., 2010) is that face stimuli are high-dimensional and have content at many levels, complicating the separation between perceptual, attentional, and executive function. Another complication is that if the observer uses only 2 response keys in a task-switching paradigm, an error could be either due to a failure to switch or to a successful switch followed by a perceptual or attentional error (Ravizza and Carter, 2008).

Here, we attempted to characterize deficits in early processes of perceptual encoding in ADHD and dissociate them from executive deficits using a visuo-motor decision-making paradigm with task-switching which avoids the complications listed above. By using a total of 8 possible buttons out of which only 2 were relevant on a given trial, our response paradigm allowed for *task-irrelevant motor output (TIMO)*, a new measure of executive control deficits. We defined a perceptual error as a press of the wrong button among the 2 relevant ones. We optimized the quantitative characterization of perceptual function by: a) using simple stimuli with feature dimensions orientation and color, thus minimizing high-level cognitive effects; b) varying stimuli parametrically along a continuum to estimate psychometric curve parameters (standard in perceptual psychophysics but still relatively rare in the study of ADHD (Friedman-Hill et al., 2010; Kim et al., 2014a,b; Roberts

et al., 2017; Stevens et al., 2012)); c) using an efficient stimulus selection method to minimize the number of trials needed for accurate estimation of parameters (Acerbi, 2016). Broadly, our work follows a recent proposal to apply four levels of analysis to computational psychiatry: development of behavioral tasks, fitting of computational models, estimating parameters, and classification for diagnosis (Wiecki et al., 2015).

## 3.2 Methods

### 3.2.1 Approach

20 ADHD and 20 Control adult participants took part in our experiment. Stimuli were colored ellipses; each display contained one stimulus on the right of the fixation dot and one on the left. The participants performed yes-no discrimination (more precisely called yes-no classification or categorization) (Ma Wei Ji, 2018). Specifically, the participants performed either fine orientation discrimination (was the cued ellipse clockwise or counterclockwise relative?) or fine color discrimination (was the cued ellipse more yellow or more blue?). The cue was 100% valid. In this task, participants had to rely on their internal memorized references, here for vertical and respectively the mid-level green in between yellow and blue.

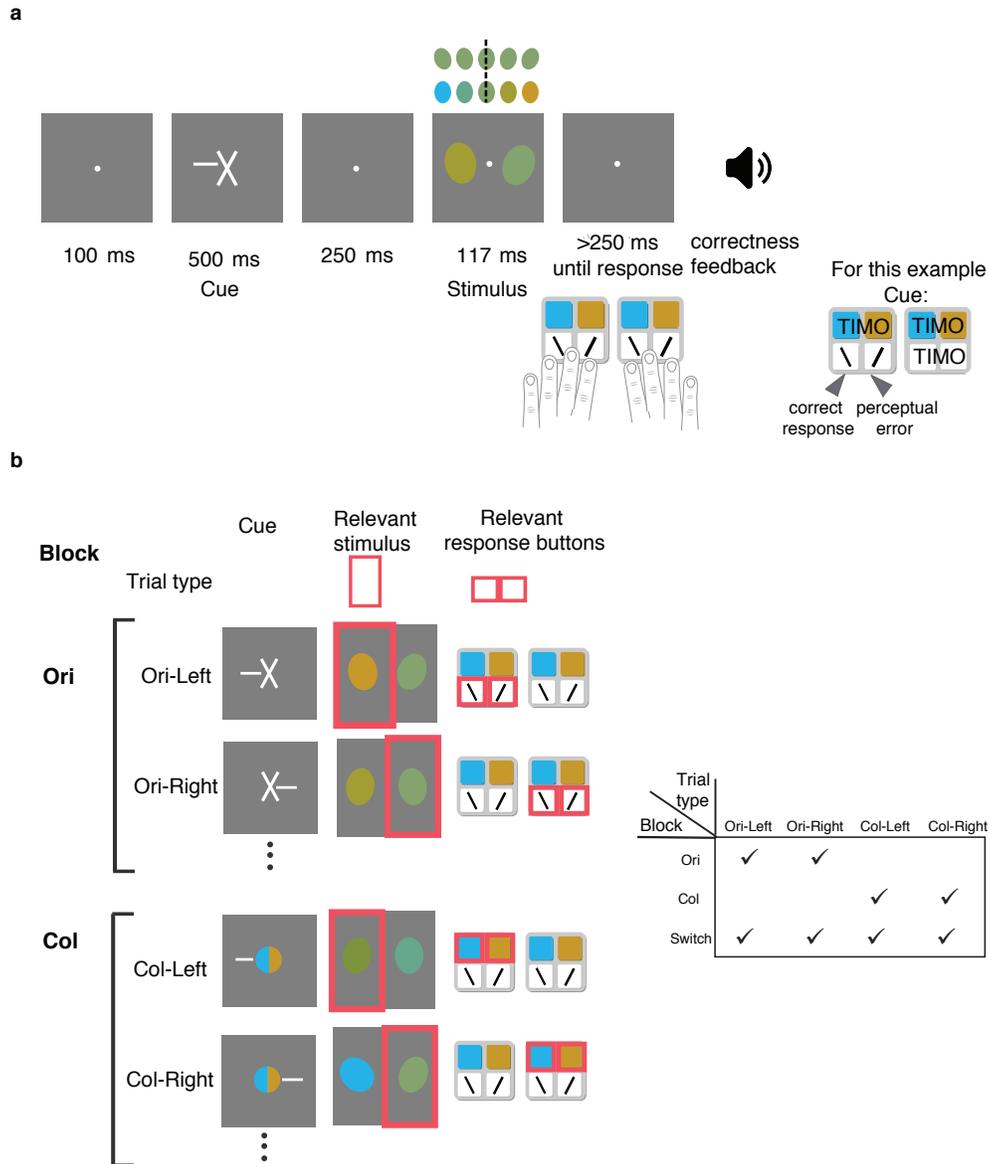
We distinguish our design from other, perhaps more common, psychophysical tasks. For instance, 2AFC discrimination tasks could ask for a comparison between the 2 presented stimuli, for instance: 'was the stimulus higher in contrast tilted to the right or to the left?' (Klein, 2001). We also distinguish our task also from orientation discrimination tasks in which, across trials, only 2 orientations are shown to be discriminated (say, -20

and +20), and other experimental manipulations are of interest.

Every trial started with a symbolic feature dimension cue, informing the participant which feature dimension was relevant on that trial. Simultaneously presented was a spatial cue (a line segment), informing the participant which side of the screen was relevant on that trial (Figure 3.1a). To better detect failures of spatial or feature switching, we used a response paradigm in which, on each trial, only 2 of 8 response keys were relevant, depending on the spatial and the feature cue; any other key press counted as task-irrelevant motor output (TIMO). Separately in each condition and for each participant, we used a Bayesian adaptive method to select maximally informative stimuli (see “Target stimulus generation”). This method allowed us to estimate the psychometric curve parameters with relatively few trials.

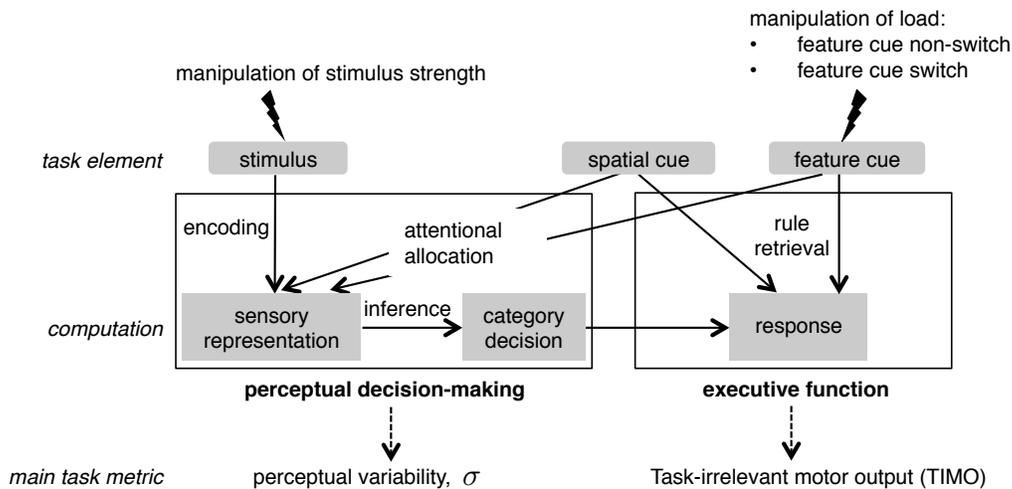
Each participant experienced three types of blocks: Ori, Col and Switch. In Ori blocks, the feature dimension cue was always orientation. The spatial cue was randomly chosen on each trial, yielding 2 possible trial types: Ori-Left and Ori-Right (Figure 3.1b). We analyzed the Ori-Left and Ori-Right trials together as the Ori condition. In Col blocks, the feature dimension cue was always color and again the spatial cue was randomly chosen on each trial, yielding 2 possible trial types, Col-Left and Col-Right, which we grouped together for analysis into the Col condition. In Switch blocks, all 4 trial types were possible. We will refer to the orientation and color trials in switch blocks as the OriS and ColS conditions, respectively, and to the difference between no-switch and switch blocks as a difference in (executive) load.

An observer’s sequence of computations in the task can be conceptualized as a perceptual decision-making stage (stimulus encoding, affected by attention, and inference), followed by executive processing (rule retrieval and response execution) (Figure 3.2). The



**Figure 3.1 Task design. (a)** Trial sequence example. A feature dimension cue indicated whether orientation (cross) - depicted here- or color (colored circle) was relevant, while a simultaneous endogenous spatial cue (line segment) indicated which side (left or right) was relevant. Thus, the participant received one of 4 possible cue screens. We always chose the spatial cue randomly. The participant had to respond whether the orientation of the ellipse on the relevant side was clockwise or counterclockwise with respect to vertical, or whether its color was more yellow or more blue, with the associated set of keys (left or right). The color and orientation continua are shown above the stimulus screen, with the dashed line at vertical and respectively mid-level green. To respond, the participant could press any one of 8 keys but only 2 were task-relevant on a given trial. The participant received correctness feedback. **(b)** (Left) Cue - relevant stimulus - relevant response buttons pairings for the 4 types of trials as they arise from the 4 feature and spatial cues combinations (2\*2). Relevant is marked with pink for visualization only. Pressing any other button would result in TIMO. (Right) During Ori and Col blocks, only 2 types of trials are possible, while during Switch blocks all 4 trial types are possible.

parametric variation of stimulus strength allowed us to estimate perceptual variability  $\sigma$  (or noise, the inverse of slope/sensitivity) as a main metric of perceptual function, and the 8-button response paradigm allows us to estimate task-irrelevant motor output as a main metric of executive function. In addition, we characterized behavior using other psychometric curve parameters, median reaction time, and reaction time variability.



**Figure 3.2 Dissociation of perceptual and executive processes.** Schematic of the perceptual and executive processes that may play a role in this task, and the corresponding task metrics.

While usually a noise parameter (equivalent to our perceptual variability) in psychometric curves reflects a mix of sensory and decision noise (Gold and Ding, 2013), we believe that here the perceptual variability parameter for orientation and color could be additionally modulated by endogenous attentional allocation. Since each ellipse was placed at 2.5 dva eccentricity, the taxing of covert endogenous attention might not have been as strong as in other studies (with more commonly used 5 dva eccentricity), but we still cannot rule out its influence. Previous studies showed modulation of psychometric curve pa-

rameters by attention, though either in different tasks such as target detection (Bashinski and Bacharach, 1980), 2AFC orientation discrimination (Downing, 1988), or color-change detection (Herman et al., 2015), or examined exogenous attention (Fuller and Carrasco, 2006), or had other stimulus strength manipulation, such as contrast (Ling and Carrasco, 2006; Pestilli et al., 2007) (for reviews see (Carrasco, 2011, 2014)).

### 3.2.2 Experimental methods

**Participants.** We recruited all participants through local advertisements, including flyers and newspaper and radio advertisements. Information on the participants is presented in Supplementary section “Demographic and clinical information”. Participants in both groups were matched as much as possible by age, sex, and education (see Table 3.2). 20 ADHD participants (12 female) of mean age 35.3 (SD: 10.0, range: 21 to 55) and 20 control participants (11 female) of mean age 32.5 (SD: 6.1, range: 19 to 44), with no statistical difference between their ages (Wilcoxon rank-sum test,  $p = 0.78$ ), participated. 17 out of the 20 ADHD participants presented the combined subtype, and 3 the inattentive subtype. All participants spoke English and had normal or corrected-to-normal vision. We asked every participant before they started if they were colorblind. One participant was excluded because of color blindness. All participants provided informed consent. The study conformed to the Declaration of Helsinki and was approved by the Institutional Review Board of New York University School of Medicine.

**Psychiatric assessment and diagnosis.** None of the participants with ADHD were prescribed or took stimulant medication within 2 months of participating in the study. Participants with comorbid anxiety or unipolar depressive disorders were included as long as the symptoms at the time of evaluation were mild or in remission. Participants with

bipolar disorders, psychotic disorders, substance use disorders, and neurologic disorders were excluded. For all adults, the diagnostic procedure included both clinician administered and self-administered scales. A trained clinician assessed every participant using the Adult ADHD Clinician Diagnostic Scale (ACDS) v.1.2, the Adult ADHD Investigator Symptom Rating Scale (AISRS), the Clinical Global Impressions-Severity of Illness (CGI-S) Scale, and the M.I.N.I International Neuropsychiatric Interview. All participants also completed the Adult ADHD Self-Report Scale (ASRS v.1.1.), the Adult ADHD Quality of Life (AAQoL) Scale, the World Health Organization Disability Assessment Schedule (WHODAS-II), and the Behavior Rating Inventory of Executive Function Adult Version (BRIEF-A). These scales have been extensively validated (Adler and Cohen, 2004; Kessler et al., 2005, 2006; Silverstein et al., 2018).

**Apparatus.** We displayed stimuli on a 23-inch (58.42 cm) Acer T232HL LCD monitor of resolution:  $1920 \times 1080$  pixels and 60 Hz refresh rate (1 frame lasting 16.7 ms). We used a Kinesis Freestyle2 split keyboard. Participants used a head rest located at approximately 55 cm from the screen; this meant that 1 degree of visual angle (dva) subtended approximately 34 pixels. Stimulus presentation and response collection were controlled by a Windows computer running Matlab 7.1 (MathWorks, Massachusetts, USA) with Psychtoolbox3 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and EyeLink (Cornelissen et al., 2002).

For 10 out of 20 ADHD participants and 10 out of 20 control participants, we monitored their fixation and recorded their eye movements. The rationale for not eye tracking all participants was a mixture of lack of sufficient time on the participants' side and balanced design on the experimenters' side. The eye tracker was calibrated using the five-point calibration routine before every block. We recorded eye movements using a remote infrared

video-oculographic system (EyeLink 1000 Plus; SR Research, Ltd, Mississauga, Ontario, Canada) with a 1 kHz sampling rate. We set the heuristic filtering option 'OFF'.

**Stimuli.** The background was mid-level gray (28.7 cd/m<sup>2</sup>). The stimuli were ellipses with area of 1600 pixels<sup>2</sup> and 0.55° eccentricity, and thus with a major axis of 50 pixels and minor axis of 41 pixels. For the non-target ellipse, the orientation was randomly drawn from a von Mises distribution centered at 0 with  $\kappa = 30$ , and then divided by 2, approximately equivalent to a Gaussian distribution with mean 0 and a standard deviation of about 5°. The color of the non-target ellipse was based on a uniformly drawn sample that was used to linearly interpolate between blue and yellow in CIE L\*a\*b\* (CIELAB) color space, with blue as [78 -30 -40], corresponding in RGB space to [0 167 255] and yellow as [78 0 80] corresponding in RGB to [200 130 0]. For each color, lightness was always kept constant at  $L = 78$ . Indeed, measured luminance was  $\sim 39$  cd/m<sup>2</sup>. The target stimulus was specified on a trial-to-trial basis using the Bayesian algorithm described below.

**Trial sequence** (Figure 3.1a). A trial sequence started with the simultaneous appearance of a feature dimension cue and a spatial cue, presented for 500 ms. The feature dimension cue for orientation consisted of 2 white line segments, each of length approximately 1 dva, crossing at the center, with orientations tilted  $\pm 26.6^\circ$  with respect to vertical; for color, it consisted of 2 semi-circles (divided vertically, right one yellow, left one blue) joined to form a circle of radius approximately 0.3 dva. Simultaneously, a spatial cue was presented, which consisted of a horizontal line segment of length approximately 0.5 dva emanating from the center of fixation to the left or to the right. We chose 500 ms to ensure sufficient time for the deployment of endogenous feature-based attention (Liu et al., 2007). Following a delay of 250 ms consisting of the presentation of a central fixation circle of radius 0.12 dva, 2 ellipses appeared at 2.5 dva to the right and left of a central fixation circle.

The stimuli were presented on the screen for 117 ms, followed by another delay period of 250 ms.

After the post-stimulus delay, the participant had to respond about the target ellipse via a specific key press out of a total of 8 keys (Figure 3.1a). On any given trial, 6 of these 8 keys are irrelevant. For orientation, the participants were instructed to press one of the 2 labeled keys for clockwise (CW) or counterclockwise (CCW), using the left keypad for the left spatial cue and the right one for the right spatial cue. For color, they had to press one of 2 labeled keys to indicate whether the ellipse was more yellow or more blue, also using the left or respectively right keypad depending on the spatial cue. Figure 3.1b shows all these 4 possible cue-response mappings. The direction of the spatial cue was randomly drawn on each trial, so participants used their right hand approximately half the time. After the response, auditory feedback was provided for 200 ms: a 1200 Hz tone if the participant had pressed the correct key, and a 500 Hz tone if the participant had pressed any of the 7 incorrect keys.

**Training.** Before they began the experiment, participants were guided step by step through the different parts of instructions. The experimenter read the instructions on the screen (presented in Figure 3.6a) out loud. To remind subjects of the stimulus-response pairings, a sheet with these pairings was posted on the wall of the psychophysics room (Figure 3.6b). In total, participants performed 40 training trials: a short orientation only block ('O') of 10 trials, a short color only block of 10 trials and a short switch block ('S') of 20 trials. The experimenter was present with the participants during the training to observe responses, provide further feedback and answer questions. Participants repeated the set of all 40 training trials until they achieved a performance greater than 65%.

**Experiment structure.** Afterwards, they performed 8 blocks of about 100 trials each

in the order 'O-C-S-S-S-S-C-O' or 'C-O-S-S-S-S-O-C', with 30 seconds breaks in between blocks. Changes in block type were signaled with a screen with the instruction 'In this block, your job is to report ORIENTATION' for O blocks, or 'In this block, your job is to report COLOR', for C, or 'In this block, your job is to report either ORIENTATION or COLOR', for S, with each feature dimension word followed by its associated symbol. In total, participants completed 800 non-aborted trials, approximately 200 in each one of the four conditions, Ori, Col, OriS and ColS (from S blocks).

### 3.2.3 Statistical analyses

For most metrics, we report median values and 95% bootstrapped confidence intervals. Across 50000 iterations, we took samples with replacement from and of the same size as the original data with Matlab's `randsample` and calculated the median of each of those sets of samples. The the 2.5th and 97.5th quantiles of the distribution of medians across iterations were taken as the 95% confidence intervals.

**Three-way mixed-design ANOVA.** To determine the differences between groups and the 2 experimental conditions of load and feature, we used three-way mixed-design ANOVA with two repeated measures, since we have one "between - participants" variable (group) and two "within - participants" factors (feature - Ori vs Col and load- No-switch vs Switch). Beforehand, we log transformed the measures that were lower bounded by 0. When we assumed shared parameters between No-switch and Switch and thus we had only one "within - participants" factor, we used two-way mixed-design ANOVA. We implemented the ANOVAs in SPSS with "General linear model: repeated measures". For post-hoc comparisons, we adjust the significance level according to the Sidak correction to  $\alpha_{\text{sid}} = 1 - (1 - \alpha)^{\frac{1}{\text{number of comparisons}}}$ . For the three-way mixed-design ANOVA, we performed,

unless otherwise specified, 12 planned pairwise comparisons in Matlab: Wilcoxon rank-sum tests between groups (one for each condition, 4 in total), and Wilcoxon signed-rank tests for conditions within a group (4 per group, 8 in total). We used the Sidak correction for multiple comparisons, decreasing the significance level to  $\alpha = 0.0043$  for post hoc comparisons following the three-way mixed-design ANOVA or respectively  $\alpha = 0.0127$  following the two-way mixed-design ANOVA.

**Pairwise correlations.** To correct for multiple comparisons when examining the pairwise correlation matrix of the performance measures, we used a method from Nyholt et al. (Nyholt, 2004). If  $M$  is the total number of measures, the number of effective comparisons will be decreased more if the measures are more highly correlated, as captured in a higher variance of the eigenvalues  $\lambda_{\text{obs}}$  of the correlation matrix, which we calculated with Matlab’s function `eig`. Then,  $M_{\text{eff}} = 1 + (M - 1) \left(1 - \frac{\text{var}(\lambda_{\text{obs}})}{M}\right)$ . As in (Nyholt, 2004),  $M_{\text{eff}}$  is used in the Sidak correction (a slightly less conservative alternative to the Bonferroni correction), modifying the significance level to  $\alpha_{\text{sid}} = 1 - (1 - \alpha)^{\frac{1}{M_{\text{eff}}}}$ .

**Linear regression.** We implemented multivariate linear regression with Matlab’s `fitlm`.

**Logistic regression for classification.** We fit the logistic regression coefficients with Matlab’s `glmfit` with input ‘binomial’ and the link parameter ‘logit’. For a given participant, we used the task metrics and the fitted coefficients with `glmval` to get  $p(\text{Diagnosis})$ , which was then thresholded at 0.5 to predict the 0 or 1 ADHD diagnosis.

**Stratified 10-fold cross-validation.** In order to assess the use of this logistic regression classifier for out-of-sample prediction, we calculated the cross-validated accuracy. We did stratified 10-fold cross-validation, in which each fold had 4 participants, 2 ADHD and 2 Controls; thus, we trained the classifier to find the coefficients over 36 participants and

tested over 4 and calculated the mean accuracy across folds. We did 1000 runs of this stratified 10-fold cross-validation to allow for different random assignments of participants into folds and took the mean accuracy over runs.

### 3.2.4 Parameter fitting

**Psychometric curves and parameters.** We fitted psychometric curves to trials on which a participant pressed one of the 2 relevant buttons.  $s$  denotes the normalized stimulus value on a given trial (ranging between  $[-0.5, 0.5]$ ). We use the following form of the psychometric curve (Wichmann and Hill, 2001):

$$p(r = 1|s; \mu, \sigma, \lambda) = \frac{1}{2} \cdot \lambda + (1 - \lambda) \cdot \Phi(s; \mu, \sigma),$$

where  $r = 1$  stands for a response “clockwise” (orientation) or for “more yellow” (color). The parameters are the point of subjective equality (PSE or bias),  $\mu$ , the inverse slope (or noise) parameter,  $\sigma$  - which both are inputs to the Gaussian cumulative density function ( $\Phi$ ) - and the lapse rate,  $\lambda$ . We had 4 conditions, Ori, OriS, Col and ColS and thus 4 psychometric curves.

**Parameter estimation and model choice.** We performed maximum-likelihood estimation of the psychometric curve parameters  $\mu$ ,  $\sigma$ , and  $\lambda$ . The likelihood of a parameter combination is the probability of the data given that parameter combination; we denote the log likelihood by LL. We assumed that trials are independent of each other and thus we summed the log likelihoods across all trials. We fitted orientation and color trials separately; thus the following log likelihoods apply to either set of trials. In the main model, we assumed that  $\mu$  and  $\lambda$  are shared across both load conditions (No-switch and Switch),

whereas  $\sigma$  might differ. These assumptions had both a practical and a principled motivation. Assuming that parameters are shared between conditions reduced the number of parameters to 8 and made parameter estimates more reliable. Moreover, if  $\mu$  reflects an overall bias and  $\lambda$  a generic lapsing process, we did not expect them to change with load. For a model without these assumptions, and a model comparison, see Supplementary section “Further information on psychometric curves and parameters”. The log likelihood for trials in a given feature dimensions becomes

$$\begin{aligned} \text{LL}(\mu, \lambda, \sigma_{\text{No-switch}}, \sigma_{\text{Switch}}) &= \log p(\text{data} \mid \mu, \lambda, \sigma_{\text{No-switch}}, \sigma_{\text{Switch}}) \\ &= \sum_{\text{No-switch trials } j} \log p(r_j \mid s_j; \mu, \lambda, \sigma_{\text{No-switch}}) + \sum_{\text{Switch trials } j} \log p(r_j \mid s_j; \mu, \lambda, \sigma_{\text{Switch}}), \end{aligned} \quad (3.1)$$

where  $s_j$  and  $r_j$  are the stimulus and the participant’s response on the  $j$ th trial, respectively. To estimate the parameters, we searched on a grid with 201 values in each dimension: for  $\mu$  linearly spaced from -0.2 to 0.2, for  $\lambda$  logarithmically spaced from 0.0001 to 0.3, and for each  $\sigma$  logarithmically spaced from 0.002 and 0.5.

**Reaction times.** For fitting ex-Gaussian distributions to reaction times, we used a custom made script modeled after an existent software package (Zandbelt, 2014).

**Data and code availability.** Clinical data is not available beyond diagnosis labels, experiment code is available upon request and behavioral data and analysis code are available at [https://github.com/lianaan/Perc\\_Var](https://github.com/lianaan/Perc_Var).

### 3.3 Results

We attempted to dissociate perceptual from executive deficits in ADHD with a new visuo-motor decision-making task with a task-switching component. This task yielded two main measures: task-irrelevant motor output and perceptual variability.

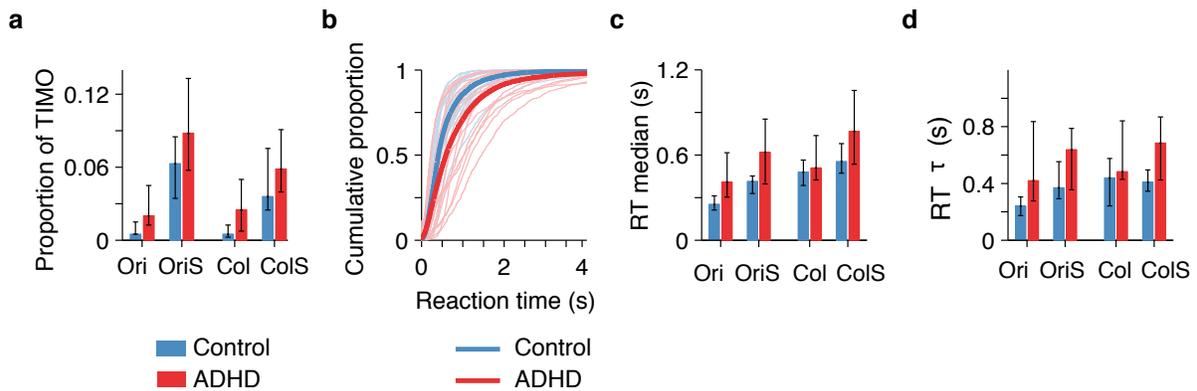
#### 3.3.1 Task-irrelevant motor output (TIMO)

TIMO refers to the trials when participants pressed one of the 6 irrelevant keys and hence such responses most likely reflect a failure of proper stimulus-response rule retrieval.

TIMO was quite low overall (mean  $\pm$  sem:  $0.06 \pm 0.01$ ); ADHD participants produced a higher proportion of TIMO ( $0.079 \pm 0.018$ ) relative to Controls ( $0.041 \pm 0.008$ ). Figure 3.3a presents a breakdown of TIMO by condition. A three-way mixed-design ANOVA on log TIMO with between-participants variable group and within-participants factors load (No-switch and Switch) and feature (Ori and Col) reveals a significant effect of group ( $F(1, 38) = 8.83, p = 0.005, \eta_p^2 = 0.19$ ), a significant effect of load ( $F(1, 38) = 101.4, p < 0.0001, \eta_p^2 = 0.73$ ), and no significant effect of feature ( $F(1, 38) = 1.62, p = 0.21, \eta_p^2 = 0.04$ ). Neither of the two-way interactions nor the three-way interaction were significant ( $p > 0.06$ ). In particular, the group  $\times$  load interaction was not significant ( $F(1, 38) = 3.72, p = 0.06, \eta_p^2 = 0.09$ ); thus, we did not find that switching between feature dimensions carries a higher cost in ADHD. Next, we performed 12 post-hoc planned comparisons: within each group, Wilcoxon signed-rank tests for Ori versus OriS, Col versus ColS, Ori versus Col, and OriS versus ColS, and between groups, Wilcoxon rank-sum tests for Ori, OriS, Col, and ColS. After Sidak correction ( $\alpha = 0.0043$ ), no between-group comparisons were significant ( $p > 0.0046$ ). The within-group load comparisons were all

significant ( $p < 0.002$ ). No within-group feature comparisons were significant ( $p > 0.07$ ). Taken together, these results validate TIMO as a metric of interest for executive control.

In the OriS and ColS conditions, the majority of TIMO seemed to be feature errors (Figure 3.7b). Relative to the instructions on a given trial, the 6 irrelevant keys subdivide into 2 that represent spatial errors, 2 feature errors and 2 that represent both spatial and feature errors. We did not delve into these distinctions since overall TIMO was quite low.



**Figure 3.3 ADHD participants had higher TIMO and longer and more variable reaction times.** (a) Proportion of TIMO across conditions. Here and elsewhere, values represent medians across participants and error bars the bootstrapped 95% confidence intervals. (b) Empirical cumulative density functions of reaction times, collapsed across all conditions. Thin lines: individual participants. Thick lines: median for the RT distribution collapsed across all participants in a group. (c) Reaction time median by condition and group. Throughout the paper, we use RT median because reaction time distributions are not Gaussian. (d) Reaction time variability metric, the  $\tau$  parameter from ex-Gaussian distribution fits, by condition and group.

### 3.3.2 Reaction times

ADHD participants showed longer reaction times (RTs; Figure 3.3b). Three-way mixed-design ANOVA on log RTs revealed a significant effect of group ( $F(1, 38) = 4.72, p = 0.036, \eta_p^2 = 0.11$ ), and significant effects of load ( $F(1, 38) = 84.92, p < 0.0001, \eta_p^2 = 0.69$ )

and feature ( $F(1, 38) = 70.23, p < 0.0001, \eta_p^2 = 0.65$ ). In addition, we found significant group  $\times$  feature ( $F(1, 38) = 4.63, p = 0.038, \eta_p^2 = 0.11$ ) and load  $\times$  feature interactions ( $F(1, 38) = 12.37, p = 0.001, \eta_p^2 = 0.25$ ), but not a significant group  $\times$  load interaction ( $F(1, 38) = 3.97, p = 0.054, \eta_p^2 = 0.095$ ). After Sidak correction ( $\alpha = 0.0043$ ), none of the between-group comparisons were significant ( $p > 0.019$ ). The effects of within-group load and feature on log RTs were all significant both within Control and within ADHD ( $p < 0.001$ ). Higher RTs for Col than Ori could be due to the fact that the Ori responses are intuitively mapped to left and right, while the Col responses are arbitrarily mapped as blue to left and yellow to right.

Higher RT variability (or intra-individual variability) in ADHD has been found consistently (Kofler et al., 2013) and has been generally thought to reflect cognitive processes separate from higher median RTs (Castellanos et al., 2006; Kofler et al., 2013) (but see (Karalunas et al., 2012) for an opposing account). The term RT variability has been used to refer to different aspects of RT distributions (Kofler et al., 2013); here, we fitted ex-Gaussian distributions (Leth-Steensen et al., 2000) and used the  $\tau$  parameter as a measure of RT variability. The  $\tau$  parameter has been shown to capture the tendency of ADHD participants to have a higher proportion of abnormally slow responses (Castellanos et al., 2006; Kofler et al., 2013; Leth-Steensen et al., 2000). Before committing to the ex-Gaussian, we verified that it captures the data better than the log-normal and gamma distributions (see Supplementary section “Further information on reaction times”). Three-way mixed-design ANOVA on log  $\tau$  revealed a significant effect of group ( $F(1, 38) = 7.72, p = 0.008, \eta_p^2 = 0.17$ ), an effect of load ( $F(1, 38) = 9.32, p = 0.004, \eta_p^2 = 0.20$ ) and an effect of feature ( $F(1, 38) = 18.85, p < 0.001, \eta_p^2 = 0.33$ ). The only significant interaction was between load and feature ( $F(1, 38) = 14.96, p < 0.001, \eta_p^2 = 0.28$ ). After Sidak correction ( $\alpha = 0.0043$ ), none of the between-group comparisons were significant

( $p > 0.006$ ). Within Controls, the effects of load and feature on log RT  $\tau$  were significant for Ori vs. OriS and Ori vs. Col ( $p < 0.001$ ). Within ADHD, no effects of load or feature were significant ( $p > 0.02$ ). We confirmed the pattern of higher RT variability in ADHD with a non-parametric measure, RT iqr (see Supplementary section “Further information on reaction times”).

Overall, we found that ADHD participants had longer and more variable reaction times, consistent with previous work (Douglas, 1999; Kofler et al., 2013; Leth-Steensen et al., 2000). However, RT-related differences are difficult to categorize as perceptual or executive because they might encompass multiple processes, including sensory encoding (possibly slower with more perceptual variability), decision time, stimulus-response rule retrieval, response preparation, and response execution (Karalunas et al., 2012). The effect of load on RT and RT  $\tau$  does seem to suggest that on Switch trials, more time is spent on executive processes, here stimulus-response rule retrieval, response preparation, and response execution, relative to No-Switch trials.

### 3.3.3 Psychometric curve parameters

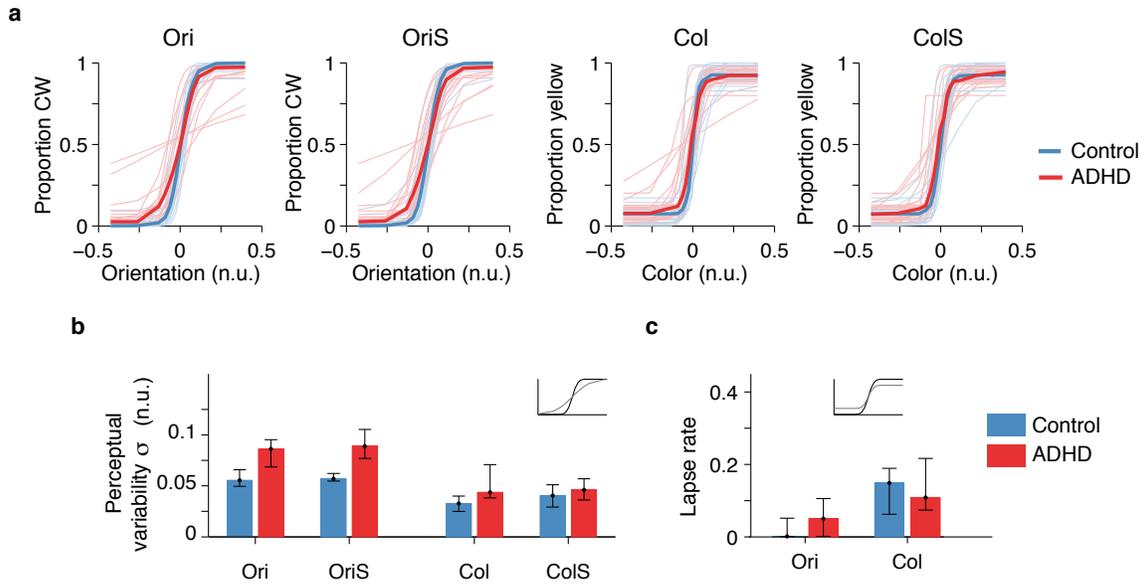
We confined the following analysis to the trials in which participants pressed one of the 2 relevant keys. Because of the Bayesian stimulus selection method, each participant received a different set of stimuli for each condition (see Figure 3.11) and thus proportion correct is largely stable across conditions and participants (mean  $\pm$  sem:  $0.811 \pm 0.007$ , Figure 3.7A) and thus not informative. Instead, we fitted a psychometric curve to non-TIMO trials in each condition (Kingdom and Prins, 2009). Thus, the parameters of the psychometric curves captured the differences in performance across conditions and participants. The normalized orientation and color continua spanned the interval  $[-0.5, 0.5]$ .

Each psychometric curve has three parameters: a point of subjective equality  $\mu$ , perceptual variability  $\sigma$ , and a lapse rate  $\lambda$  (Figure 3.4b,c and Figure 3.13d). Non-zero  $\mu$  represents a tendency to choose one option more than the other. The parameter  $\sigma$  is a composite of sensory noise and decision noise, and also reflects the quality of the allocation of spatial attention, and of feature attention in switch blocks. Higher  $\sigma$  denotes a reduced ability to discriminate between small variations within a feature. The parameter  $\lambda$  reflects trials with lapses in attention or erroneous motor output. In our main model, we assumed that  $\mu$  and  $\lambda$  are independent of load; we confirmed this assumption by comparing to a model without these assumptions (“full” model) in the Supplementary section “Further information on psychometric curves and parameters”. The parameters  $\sigma$  and  $\lambda$  might trade off against each other, although this is less of a concern in our main model than in the “full” model.

Three-way mixed-design ANOVA on  $\log \sigma$  showed a significant effect of group ( $F(1, 38) = 10.56, p = 0.002, \eta_p^2 = 0.22$ ), a significant effect of feature ( $F(1, 38) = 37.2, p < 0.001, \eta_p^2 = 0.49$ ), but not of load ( $F(1, 38) = 0.97, p = 0.33, \eta_p^2 = 0.025$ ). The effect of group  $\times$  load was not significant ( $F(1, 38) = 2.0, p = 0.16, \eta_p^2 = 0.05$ ). Because the normalization to the (arbitrary) stimulus range is specific to each feature dimension, the values of  $\sigma$  cannot be meaningfully compared between orientation and color: a different stimulus range would have changed the  $\sigma$  values without changing the observer’s true perceptual variability. Therefore, only the within-feature post-hoc comparisons are meaningful, giving a corrected significance level of  $\alpha = 0.0065$ . Then, the between-group comparisons were significant for both Ori and OriS ( $p < 0.0005$ ), but not for Col ( $p = 0.0083$ ) or ColS ( $p = 0.28$ ). No post-hoc comparisons with load were significant neither within Controls nor within ADHD ( $p > 0.01$ ). Higher  $\sigma$  for orientation in ADHD could result from worse low-level sensory encoding (e.g. higher neural noise), lower covert endogenous attention, higher

decision noise, or even noise in the inference process about the perceptual category. The lapse rate reflects responses that are independent of the stimulus, such as lapses of attention, but could also trade off with the  $\sigma$  parameter. Two-way mixed-design ANOVA on  $\log \lambda$  showed a large effect of feature ( $F(1, 38) = 28.08, p < 0.0001, \eta_p^2 = 0.43$ ), but no significant effect of group ( $F(1, 38) = 1.72, p = 0.19, \eta_p^2 = 0.04$ ) and no significant group  $\times$  feature interaction. After Sidak correction ( $\alpha = 0.0127$ ), we found that Control ( $p < 0.0001$ ) and ADHD ( $p = 0.011$ ) participants tended to lapse more for color than for orientation, possibly because the stimulus-response mapping is less intuitive. Results for  $\mu$  are in the Supplementary section “Further information on psychometric curves and parameters”. In conclusion, the parametric variation of low-level stimulus variables combined with stimulus optimization revealed robust perceptual deficits in ADHD, especially for orientation.

A possible cause of the increased perceptual variability in ADHD could be that ADHD participants were slower to learn the task. To check for learning, we fitted two psychometric curves for each condition, one to the first half of the trials and one to the second half. The  $\sigma$  parameters across participants, conditions and time are presented in Figure 3.14. Visually, we notice a slight improvement in perceptual variability in the second half of the trials (Figure 3.14). To quantify it, we performed a 4-way mixed-design ANOVA on  $\log \sigma$  with time as an additional factor. We found an effect of time ( $F(1, 38) = 12.7, p = 0.001, \eta_p^2 = 0.25$ ). We do not find a significant group  $\times$  time interaction ( $F(1, 38) = 0.42, p = 0.52, \eta_p^2 = 0.01$ ) and thus we have no evidence for a differential learning pattern for ADHD relative to Controls.



**Figure 3.4 Fitted psychometric curves and parameters. ADHD participants had higher perceptual variability. (a)** Psychometric curve fits across all conditions. Here and elsewhere, n.u. stands for normalized units. Thin lines: individual participants. Thick lines: medians for each group. For fits overlaid on top of data, see Figure 3.13. **(b)** Perceptual variability parameter values, medians and bootstrapped 95% confidence intervals. Top inset plot: black psychometric curve has low noise, while the grey has higher noise. **(c)** Lapse rate. Top inset plot: black psychometric curve has low lapse, while the grey has higher lapse.

**Table 3.1** Pairwise Spearman correlations across log task metrics (both behavioral and clinical). Both TIMO and perceptual variability are significantly correlated with several other variables. Boldfaced denotes significance after multiple-comparisons correction ( $\alpha = 0.0089$ , see Methods).

	TIMO	RT	RT $\tau$	Perceptual variability ( $\sigma$ )	Lapse rate ( $\lambda$ )	GEC
TIMO						
RT	$\rho = \mathbf{0.46}$ $p = 0.003$					
RT $\tau$	$\rho = \mathbf{0.42}$ $p = 0.007$	$\rho = \mathbf{0.84}$ $p < 0.0001$				
Perceptual variability ( $\sigma$ )	$\rho = \mathbf{0.41}$ $p = 0.0085$	$\rho = \mathbf{0.55}$ $p = 0.0003$	$\rho = \mathbf{0.57}$ $p = 0.0002$			
Lapse rate ( $\lambda$ )	$\rho = \mathbf{0.46}$ $p = 0.003$	$\rho = 0.23$ $p = 0.15$	$\rho = 0.17$ $p = 0.30$	$\rho = 0.28$ $p = 0.08$		
GEC	$\rho = \mathbf{0.53}$ $p = 0.0005$	$\rho = 0.25$ $p = 0.12$	$\rho = 0.34$ $p = 0.03$	$\rho = \mathbf{0.50}$ $p = 0.0009$	$\rho = 0.30$ $p = 0.06$	
ACDS	$\rho = 0.40$ $p = 0.01$	$\rho = 0.31$ $p = 0.05$	$\rho = \mathbf{0.45}$ $p = 0.004$	$\rho = \mathbf{0.51}$ $p = 0.0008$	$\rho = 0.18$ $p = 0.26$	$\rho = \mathbf{0.80}$ $p < 0.0001$

### 3.3.4 Correlations across metrics

Next, we asked whether behavioral metrics are correlated with each other (Table 3.1). For this analysis, we collapsed across groups; per participant, we averaged each behavioral metric across all four conditions. We found that the perceptual variability parameter  $\sigma$  is significantly correlated with TIMO, RT median, and RT  $\tau$ , with high effect sizes. Note that the perceptual variability parameter and TIMO were computed from different sets of trials, therefore reducing the probability that their correlation is spurious. In addition, a breakdown of some of these correlations by group, symptom type, and condition is presented in Supplementary section “Breakdown of correlations”.

So far, we have characterized behavioral differences between ADHD and Controls in our task. Next, we asked if behavioral metrics relate to common clinical ones, namely the General Executive Composite score (GEC), as assessed by the Brief-A questionnaire (self-reported, (Roth et al., 2013)), as well as the (ACDS) scores (clinician interview, (Adler and Cohen, 2004)). The GEC and ACDS (Table 3.3) are meant to be continuous measures

of executive control and symptom severity, respectively. Both GEC and ACDS revealed strong correlations with TIMO, suggesting that TIMO could serve as a behavioral marker of executive deficits. GEC and ACDS were both also strongly correlated with perceptual variability. In addition, ACDS (but not GEC) was correlated with RT  $\tau$ , which provides a graded counterpart of the robust finding of increased RT variability in ADHD (Kofler et al., 2013). A linear regression of GEC as a function of behavioral metrics ( $R^2 = 0.38$ ) showed only TIMO as statistically significant (Table 3.8a), reinforcing our interpretation of TIMO as reflecting failures of executive function. A linear regression of ACDS as a function of the same metrics ( $R^2 = 0.33$ ) only showed significance for RT  $\tau$  (Table 3.8b), suggesting that GEC and ACDS, despite being strongly correlated (Figure 3.16), could capture distinct aspects of impairment (Adler et al., 2017). However, the determinant of the correlation matrix of these measures is 0.11, nearing 0 and thus signaling multicollinearity (Dormann et al., 2012). Therefore, we have to be cautious in interpreting the individual contributions of these regressors. Nevertheless, these results suggest that our behavioral metrics capture to some extent the same processes as clinical metrics, while having the advantage of avoiding the potential subjectivity inherent in questionnaires.

### 3.3.5 Classification of participants

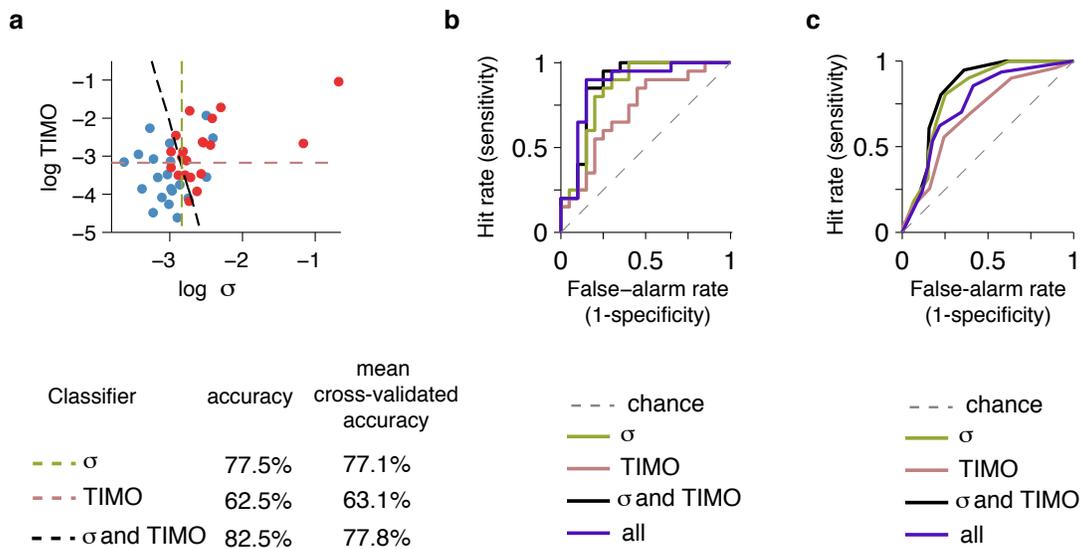
Finally, we asked how accurately we can classify a given participant as ADHD or Control based purely on behavioral task metrics. Figure 3.5 depicts these results. A logistic regression using only the perceptual variability parameter yielded a classification accuracy of 77.5%, with a hit rate (sensitivity) of 75% and a false-alarm rate (1 minus specificity) of 20% (Figure 3.5a). A logistic regression classifier based on TIMO only had an accuracy of 62.5%, with a hit rate of 65% and a false-alarm rate of 40%; using both percep-

tual variability and TIMO improved the accuracy to 82.5%, with a hit rate of 80% and a false-alarm rate of 15% (Figure 3.5a). Of note, while these variables are correlated, the determinant of their correlation matrix is 0.82, far enough from 0 that collinearity should not be a problem (Dormann et al., 2012). Adding more regressors (RT, RT  $\tau$  and lapse) did not yield further improvement (80.0%) ; this is not surprising in light of multicollinearity. Hence, we consider perceptual variability and TIMO as the main regressors of interest. In order to assess the use of this logistic classification for out-of-sample prediction, we did stratified 10-fold cross-validated and found mean accuracies of 77.1% with perceptual variability as the only regressor, 63.1% with TIMO only, 77.8% with both perceptual variability and TIMO and 70.0% with all metrics. The relatively high classification performance suggests that our task has potential as a diagnostic tool.

In addition to thresholding at 0.5 to get diagnosis and subsequently accuracy as above, we also thresholded  $p(\text{Diagnosis})$  at linearly spaced values between 0 and 1 and plotted the resulting receiver operating characteristic (ROC) curves, both without (Figure 3.5b) and with stratified 10-fold cross-validation (Figure 3.5c). As expected, the ROC curve for the classifier all metrics shows the best performance (highest area under the curve (AUC)) without cross-validation, but its performance degrades for out-of-sample predictions in the cross-validated case.

### 3.4 Discussion

In this study, we dissociated stimulus encoding (perceptual) from response selection (executive) deficits in ADHD with a new visuo-motor decision-making task with a task-switching component. To better separate executive deficits from perceptual and at-



**Figure 3.5** Logistic regression based on task metrics can classify participants into ADHD and Controls with accuracies larger than 70%. **(a)** Dots: combinations of log TIMO and log perceptual variability ( $\sigma$ ) across participants. Dashed lines: logistic regression classifiers trained on log  $\sigma$  only (olive), TIMO only (old rose) and both (black). **(b)** Full ROC curves obtained by varying the diagnosis threshold for the three classifiers in **(a)**, as well as for one based on all 5 behavioral metrics (purple). **(c)** Full ROC curves, this time with stratified 10-fold cross-validation, for the same classifiers as in **(b)**.

tentional failures, we used 8 response keys, 6 of which were irrelevant on any given trial (TIMO). To assess perceptual precision, we used simple stimuli that varied continuously along one dimension. We used and an adaptive stimulus selection method (Acerbi, 2016) that reduced the number of trials needed for accurate parameter estimation (relative to, for instance, the method of constant stimuli); reducing the number of trials is crucial when running the ADHD population. We found differences between ADHD and Controls in our main task metrics, TIMO (Figure 3.3a) and perceptual variability (Figure 3.4b), as well as median reaction times and reaction time variability (Figure 3.3c and d). We found correlations of these behavioral metrics with clinical metrics (Table 3.1) and were able to classify participants into ADHD and Controls with high accuracy solely on the basis of our main behavioral metrics (Figure 3.5).

Our finding of higher TIMO in ADHD could be due to more spatial switching errors or more feature switching errors, but it is hard to quantify these contributions since TIMO was overall relatively low. It is conceivable that a less intuitive stimulus-response mapping for orientation discrimination (stimulus oriented towards left/respond with key on the left), or types of stimuli that require spatial integration (Greenberg et al., 2010; Liu, 2003; Mante et al., 2013; Siegel et al., 2015) or cross-modal switching (Haigh et al., 2016), or more complex forms of task switching (Boonstra et al., 2005; Willcutt et al., 2005), would produce larger differences on a TIMO-like executive function measure.

In line with previous work (Douglas, 1999; Kofler et al., 2013; Leth-Steensen et al., 2000), we found that ADHD participants had longer and more variable reaction times. While accuracy was maintained to be approximately stable in all participants, perceptual variability was higher in ADHD, and thus the increased reaction times are not reflective of speed-accuracy trade-offs. In addition, our paradigm allowed us to analyze correlations

across individuals between reaction time metrics and other metrics. The correlation between the perceptual variability parameter  $\sigma$  and median reaction time is consistent with a drift-diffusion model, in which slower accumulation of evidence simultaneously leads to lower accuracy and longer reaction times. Indeed, many studies have found slower drift rates in ADHD (Huang-Pollock et al., 2016; Karalunas and Huang-Pollock, 2013; Karalunas et al., 2012; Lúcio et al., 2016; Ziegler et al., 2016).

We found higher  $\sigma$  in ADHD than in controls. This parameter - which we called the perceptual variability parameter could be affected both by sensory encoding (affected by attention) and decision processes. Could the differences in  $\sigma$  be attributed to either type of process? Sensory and decision noise are usually confounded in the parameters derived from behavior in common discrimination tasks (Gold and Ding, 2013). However, tasks exist in which the influences of sensory and decision noise can be separated (Drugowitsch et al., 2016; Keshvari et al., 2012). Additionally, neural data with high temporal resolution such as EEG or MEG could separate perceptual from decision-related variability as early vs late activity relative to stimulus onset (Gonen-Yaacovi et al., 2016; Mostert et al., 2015). Decision noise in perceptual decision-making might be related to decision noise on action selection in reinforcement learning models of high-level cognitive tasks. (Hauser et al., 2014) found increased decision noise (temperature parameter) in ADHD and later proposed this to underlie behavioral variability found in ADHD more generally (Hauser et al., 2016). Our result of increased perceptual variability parameter in ADHD is consistent with this general proposal, and extends it to include the possibility of an even lower-level correlate of behavioral variability.

Earlier studies examining perceptual function in isolation did not find differences between ADHD and Controls (see (Fuermaier et al., 2017) for a review). Our result of

higher perceptual variability in the ADHD group suggests that the encoding of visual stimuli is less precise than in Controls, at least when experimental conditions simultaneously tax other processes. In our case, participants had to allocate either spatial attention or both spatial and feature-based attention, as well as employ executive function by maintaining and acting on either 2 (no-switch) or 4 (switch) stimulus-response rules. Earlier studies examining covert spatial attention while attempting to minimize executive load did not find differences between ADHD and Controls (Cubillo et al., 2010; Huang-Pollock and Nigg, 2003; Roberts et al., 2017; Rubia et al., 2010). While perceptual precision and attention might be comparable between ADHD and Controls when studied in isolation, it is possible that asking ADHD participants to simultaneously devote brain resources to other processes might allow for differences in perceptual variability to emerge.

### **3.4.1 Possible lower-level neural correlates of behavioral variability in ADHD**

Our results could speak to the question of low-level perceptual components interacting with measured executive control deficits, as we found a significant correlation between the perceptual variability parameter and the executive control metric TIMO. In particular, our results raise the possibility of a shared neural source of perceptual and executive function deficits, such as a lower signal-to-noise ratio in early brain areas. While ideas of lowered signal-to-noise ratio implemented through impaired dopamine and noradrenaline signaling in ADHD have been put forward before, they have been mainly confined to cerebellar, striatal and prefrontal regions (del Campo et al., 2011; Frank et al., 2006; Hauser et al., 2016). Beyond that, one study found higher neural noise in the visual and auditory cortex of ADHD participants (Gonen-Yaacovi et al., 2016). ADHD participants could have higher perceptual variability in orientation by having less selective orientation tuning of

cells in V1; this was the mechanism proposed to underlie decreased orientation discrimination with aging in monkeys (Leventhal et al., 2000). The list of regions with lower signal-to-noise ratio in ADHD could include deeper brain structures with roles in selecting relevant sensory stimuli and maintaining stimulus-response rule representations such as the thalamus (Halassa and Kastner, 2017; Schmitt et al., 2017; Wells et al., 2016; Wimmer et al., 2015; Young and Wimmer, 2017), or even lower regions with roles in orienting of attention and behavioral flexibility, such as the superior colliculus (Krauzlis et al., 2013; Overton, 2008) or the locus coeruleus (Aston-Jones et al., 1999; Devilbiss and Berridge, 2006). However, these regions do not just modulate cortical representations but also receive substantial top-down inputs, so the source of the reduced signal-to-noise ratio could originate from either lower or higher-level brain regions.

Based on our data, we cannot establish whether the proposed low-level correlate of behavioral variability is reflective of a diffuse deficit, of frontal-based executive function, or of impairments in endogenous attention reliant on fronto-parietal circuits. Nevertheless, our results make the case that low-level perceptual function in ADHD deserves further investigation and that future task designs can easily include assessments of perceptual function in conjunction with attention and executive function. Using simple rather than high-level cognitive stimuli has the advantage that they can be used in parallel human and animal studies. Studies on mouse (Leo and Gainetdinov, 2013; Majdak et al., 2016) and rat (Clements et al., 2014) models of ADHD will provide further insight into the neural circuits implicated in ADHD and how medications can alter these circuits (Hetherington et al., 2017; Mueller et al., 2017).

### 3.4.2 Perceptual variability as a candidate diagnosis marker for ADHD

ADHD diagnosis still relies predominantly on self and sometimes collateral reports and widely accepted “psychomarkers” (also called “neurocognitive endophenotypes”) and biomarkers are lacking (Thome et al., 2012). For our findings to have implications for clinical practice, it is necessary that our task metrics are predictive of clinical metrics. We found that this was indeed the case. First, based on perceptual variability alone, we were able to classify participants into ADHD and Control with cross-validated mean accuracy of 77.0% (including TIMO, 77.7%). Beyond binary classification, we also found strong correlations between behavioral metrics ( $\sigma$ , TIMO, and RT  $\tau$ ) and clinical ones (GEC and ACDS). Based on these correlations, the behavioral metrics in our task could be considered candidate psychomarkers for ADHD, similar to the performance on the CPT (Ogundele et al., 2011), response variability (Castellanos and Tannock, 2002; Henrandez-Henrandez et al., 2015), and drift rate (Salum et al., 2014), and along with potential biomarkers such as saccade patterns (Munoz et al., 2003), microsaccade rate in specific tasks (Dankner et al., 2017; Fried et al., 2014; Panagiotidi et al., 2017), pupil size (Wainstein et al., 2017), eye vergence (Casal et al., 2018) and cortical thickness (Mous et al., 2014). While there is a long pipeline from task and metric to clinically useful assay (Hitchcock et al., 2017; Paulus et al., 2016), simple behavioral paradigms and modeling applied to ADHD and other disorders could in the long term help refine diagnostic categories and inform and quantify the efficacy of treatment, as is the goal in computational psychiatry more broadly (Montague et al., 2012; Redish and Gordon, 2016; Wiecki et al., 2015).

## 3.5 Supplementary

### 3.5.1 Demographic and clinical information

	Demographics		
Group	Gender	Age	# White
Control	11F, 9M	32.5 ± 6.1	11
ADHD	12F, 8M	35.3 ± 10.0	9

**Table 3.2 Demographic information of participants.**

Values represent mean and standard deviation.

	ADHD scale scores					
Group	ACDS	ACDS2	ASRS	AISRS	MCI	GEC
Control	25.1 ± 4.1	1.3 ± 1.7	19.3 ± 8.8	6.9 ± 4.7	45.7 ± 8.1	45.6 ± 8.5
ADHD	52.8 ± 6.8	14.4 ± 2.5	49.7 ± 6.6	36.5 ± 7.9	73.7 ± 9.0	71.2 ± 6.8
Wilcoxon rank-sum <i>p</i> values	< 10 <sup>-7</sup>	< 10 <sup>-7</sup>	< 10 <sup>-5</sup>	< 10 <sup>-6</sup>	< 10 <sup>-6</sup>	< 10 <sup>-6</sup>

**Table 3.3 Psychiatric scores of participants.**

Values represent mean and standard deviation. ACDS denotes ACDS B1-B18, and ACDS2 B22-B39.

	ASRS	AISRS	ACDS	ACDS2	GEC
ASRS					
AISRS	0.85***				
ACDS	0.90***	0.96***			
ACDS2	0.85***	0.96***	0.94***		
GEC	0.85***	0.77***	0.82***	0.81***	
MCI	0.81***	0.81***	0.84***	0.83***	0.95***

**Table 3.4 Spearman correlations across the scores for all diagnosis scales.**

10 participants were excluded from this table because not all records were available.

However, ACDS, ACDS2, GEC and MCI were available for all participants.

a

Instructions for Orientation only blocks

On every trial, you will see 2 ellipses, each with a color and an orientation.



In this block of the experiment, your job is to report ORIENTATION.

On each trial, BEFORE the ellipses, you will see X- to remind you that you have to report orientation.

You notice a short horizontal line next to the big X above.

If it points RIGHT, report on the RIGHT ellipse with the RIGHT keyboard.  
If it points LEFT, report on the LEFT ellipses with the LEFT keyboard.

If the ellipse is tilted to the left, press the button with \ or right, press the button /.

Let us do 10 trials now.  
You will get feedback: up tone if correct, down tone if incorrect.  
*Press any key to start*

Instructions for Color only blocks

On every trial, you will see 2 ellipses, each with a color and an orientation.



In this block of the experiment, your job is to report COLOR.

On each trial, BEFORE the ellipses, you will see - to remind you that you have to report color.



You notice a short horizontal line next to the big symbol above.

If it points RIGHT, report on the RIGHT ellipse with the RIGHT keyboard.  
If it points LEFT, report on the LEFT ellipses with the LEFT keyboard.

If the ellipse is more blue, press the button with or more yellow, press the button .

Let us do 10 trials now.  
You will get feedback: up tone if correct, down tone if incorrect.  
*Press any key to start*

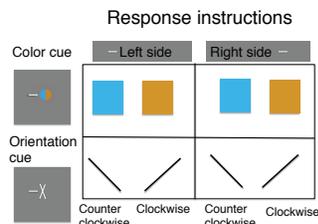
Instructions for Switch blocks

In this block your job is to report either ORIENTATION X- or COLOR - , as indicated by the symbol at the beginning of each trial.

In other words, what you have to report may SWITCH from trial to trial.

Use the same keys as before to respond.  
Let us try this for 20 trials. As before, you will get feedback.  
*Press any key to start*

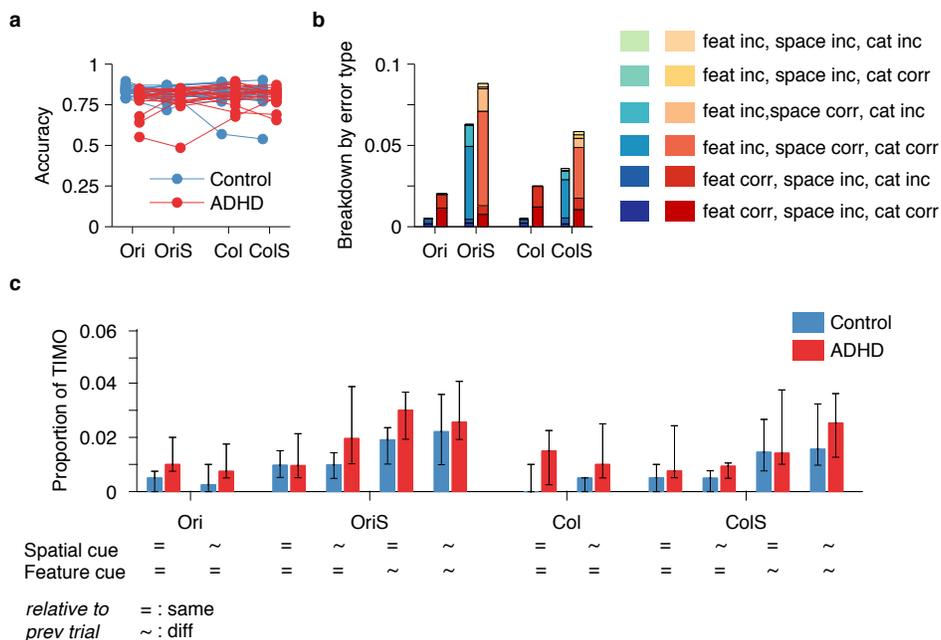
b



**Figure 3.6 Training information.** (a) Training instructions for the 3 different types of blocks: Orientation only, Color only and Switch. (b) Reminder of the stimulus-response pairings. A sheet containing this information was present on the wall of the psychophysics room within participants' sight.

### 3.5.2 Further characterization of responses

Accuracy was maintained approximately constant across participants and conditions (mean  $\pm$  sem:  $0.811 \pm 0.007$ ) due to the Psybayes method of adaptive thresholding (Figure 3.7a). We further characterized the TIMO responses, first with a breakdown by error type, available for 32 participants (Figure 3.7b) and then according to the type of the previous trial (Figure 3.7c). In the Switch trials, the majority of TIMO seem to be feature errors, possibly because the mapping from left/right visual field to left/right keyboard was more intuitive than from feature dimension to top/down of a keyboard (Figure 3.7b). Lastly, we saw no clear pattern from breaking down proportion of TIMO by the type of the previous trial (Figure 3.7c).

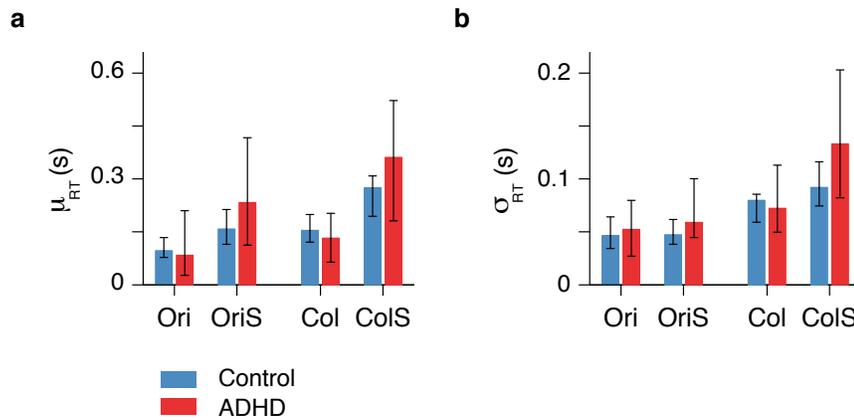


**Figure 3.7 Further characterization of responses.** (a) Accuracy as proportion correct on the trials when participants selected one of the 2 relevant keys. (b) The task irrelevant motor output from Figure 3.1, broken down by error type. For every condition, the first bar is Controls, and the second one ADHD. (c) Same, broken down by the type of the previous trial.

### 3.5.3 Further information on reaction times

#### ex-Gaussian model

ex-Gaussian distributions are commonly fitted to reaction time data and are defined by adding 2 random variables, a Gaussian with parameters  $\mu$  and  $\sigma$  and an exponential with parameter  $\tau$ . While in our data  $\tau$  showed an effect of group, neither  $\log \mu_{\text{RT}}$  ( $F(1, 38) = 0.05, p = 0.83, \eta_p^2 = 0.001$ ) nor  $\log \sigma_{\text{RT}}$  ( $F(1, 38) = 0.27, p = 0.61, \eta_p^2 = 0.007$ ) did, consistent with other studies that showed significant effects of group on  $\tau$  but not on  $\mu_{\text{RT}}$  (see (Kofler et al., 2013) for a meta-analysis).

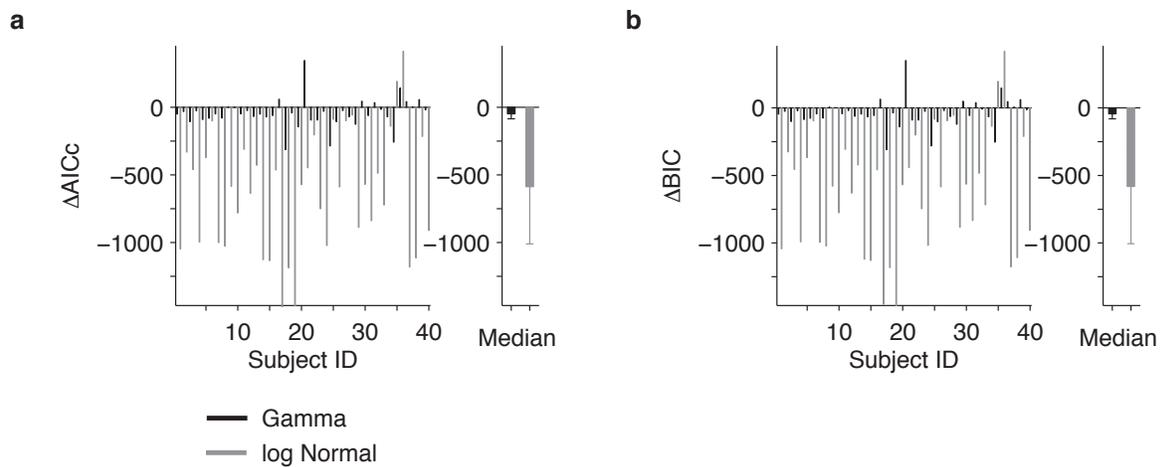


**Figure 3.8** ex-Gaussian parameters fitted to the reaction time distributions across conditions and groups. **(a)** Gaussian mean  $\mu_{\text{RT}}$ , **(b)** Gaussian standard deviation  $\sigma_{\text{RT}}$ , both for each task condition.

#### Alternative models

While ex-Gaussian distributions are routinely used to fit reaction times, they are rarely compared to alternative distributions. We used the corrected Akaike Information Criterion (AICc) and the Bayesian Information criterion (BIC) to compare the ex-Gaussian fits with

the fits of 2 other distributions on the positive real line: log-Normal and Gamma. These metrics are defined as  $AICc = -2LL^* + 2k + \frac{2k(k+1)}{n_{\text{trials}} - k - 1}$  (Hurvich and Tsai, 1989) and  $BIC = -2LL^* + k \log n_{\text{trials}}$  (Schwarz, 1978), respectively, where  $LL^*$  is the maximum log likelihood,  $k$  is the number of free parameters, and  $n_{\text{trials}}$  is the number of trials. We found that indeed the ex-Gaussian distribution was a better fit than both the log-Normal (in median by 611 according to AICc and by 607 according to BIC) and the Gamma distribution (in median by 50 according to AICc and by 45 according to BIC); see Figure 3.9 for individual subjects.

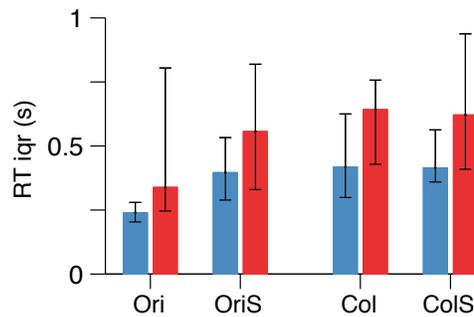


**Figure 3.9** Model comparison justifies the parametrization of reaction times with the ex-Gaussian distribution. **(a)** The ex-Gaussian model has the lowest AICc across the population **(Right)** and for almost all individual subjects **(Left)**. **(b)** Same result for BIC.

### Non-parametric measure of RT variability

We complemented the results about RT  $\tau$  (Figure 3.3) with a non-parametric robust measure of intra-individual reaction time variability, the reaction time inter-quartile range (iqr) (Figure 3.10). Three-way mixed-design ANOVA on log RT iqr's revealed a significant effect of group ( $F(1, 38) = 5.13, p = 0.029, \eta_p^2 = 0.12$ ), load ( $F(1, 38) = 18.84, p <$

0.001,  $\eta_p^2 = 0.33$ ), feature ( $F(1, 38) = 21.38, p < 0.001, \eta_p^2 = 0.36$ ), and a significant load  $\times$  feature interaction ( $F(1, 38) = 22.12, p < 0.001, \eta_p^2 = 0.37$ ). No other two-way interaction nor the three-way interaction were significant ( $p > 0.36$ ). After Sidak correction ( $\alpha = 0.0043$ ), none of the between-groups comparisons were significant. Within Controls, the effects of load and feature on log RT iqr were significant for Ori vs OriS ( $p = 0.0015$ ) and Ori vs Col ( $p < 0.001$ ); within ADHD, the only significant effect was of load for Ori vs OriS ( $p = 0.002$ ).



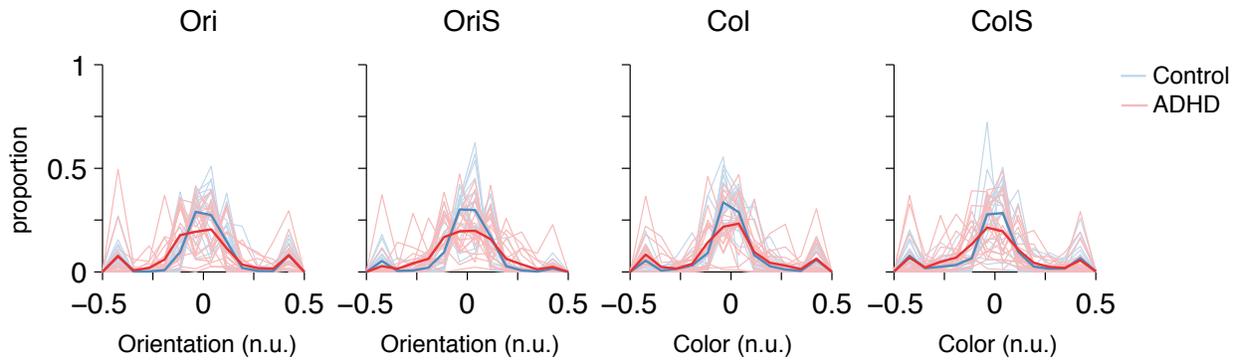
**Figure 3.10** Reaction time variability is higher in ADHD also according to a non-parametric metric, RT iqr.

### 3.5.4 Further information on psychometric curves and parameters

#### Stimuli sets

Figure 3.11 depicts the histograms of selected stimuli for each condition and each participant, optimized with the Bayesian stimulus selection method. As a consequence of this method, proportion correct is largely stable across conditions and participants (see Results 3.3.3), and the differences between participants were quantified through the psychometric curve parameters. In line with ADHD participants having higher perceptual variability, we see here that the collapsed histograms across all participants within a group

show that Controls received a higher proportion of more difficult stimuli (higher bump around 0).



**Figure 3.11 Distributions of stimuli across conditions and participants.** Thin lines: individual participants. Thick lines: proportion of stimuli collapsed across all participants within a group.

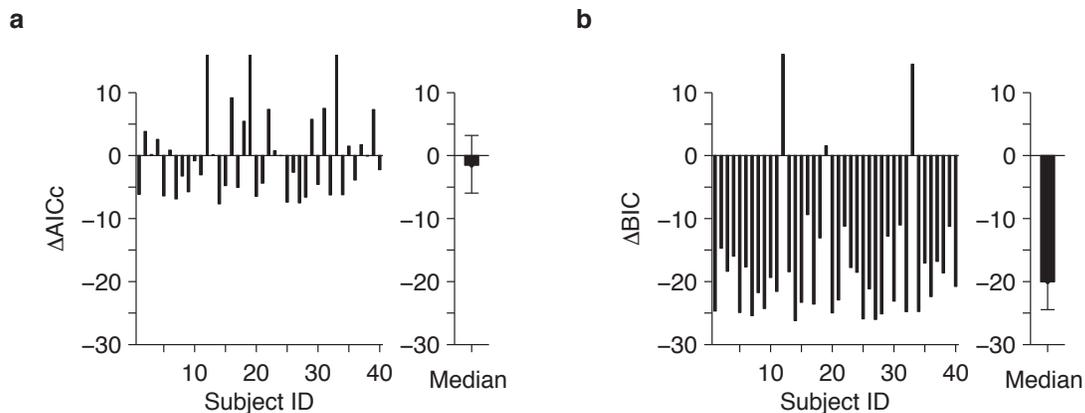
## PSE

Figure 3.13d shows the estimates of  $\mu$  (PSE) in the “shared” (main) model. Two-way mixed-design ANOVA on  $\mu$  with within-group factor feature showed an effect of group ( $F(1, 38) = 9.47, p = 0.004, \eta_p^2 = 0.2$ ), but no significant effect of feature ( $F(1, 38) = 1.17, p = 0.28, \eta_p^2 = 0.03$ ) and not a significant interaction. After Sidak correction ( $\alpha = 0.0253$ ), no effects were significant. We chose to interpolate color values between blue and yellow since the S-cone pathway is of special interest in ADHD (Tannock and Banaschewski, 2006). While we found an overall group effect on  $\mu$ , after Sidak correction the post hoc effect for color failed to reach significance, thus making our results at this point inconclusive about whether ADHD participants have different S-cone dependent color processing.

### “Full” model, 12 parameters

While in the main or “shared” model with 8 parameters (Figure 3.4) we assumed that  $\mu$  and  $\lambda$  were shared within a feature across load conditions, in the “full” model we did not constrain any parameters, yielding 12 parameters total.

As expected, the “full” model captured the data at least as well as the “shared” model. However, the “shared” model provided either a comparable (in median better by -1.5 according to AICc) or better (in median by -20 according to BIC) than the “full” model (Figure 3.12). This confirmed the plausibility of the shared-parameters assumption in the main model.



**Figure 3.12 Model comparison justifies using the “shared” model. (a)** AICc of the “shared” model minus AICc of the “full” model for **(Left)** individual subjects and **(Right)** Group - median and 95% bootstrapped confidence intervals. **(b)** Same for BIC.

We also performed three-way mixed-design ANOVA on the parameter estimates from the “full” model (Figure 3.13B). Just like in the “shared” model, we found a significant effect of group for log perceptual variability ( $\sigma$ ) ( $F(1, 38) = 5.21, p = 0.028, \eta_p^2 = 0.12$ ), a significant effect of feature ( $F(1, 38) = 37.11, p < 0.0001, \eta_p^2 = 0.49$ ), but no significant effect of load ( $F(1, 38) = 0.03, p = 0.87, \eta_p^2 = 0.001$ ). Neither of the two-way interactions nor

the three-way interaction were significant ( $p > 0.06$ ). After Sidak correction ( $\alpha = 0.0065$ , 8 comparisons, since, as in the main model, we excluded across feature comparisons due to their different units ) we found a between-group effect for Ori with  $p < 0.0001$  and OriS ( $p < 0.0025$ ), but no significant effects of group for neither Col ( $p = 0.0063$ ) nor ColS ( $p = 0.47$ ) (Figure 3.13).

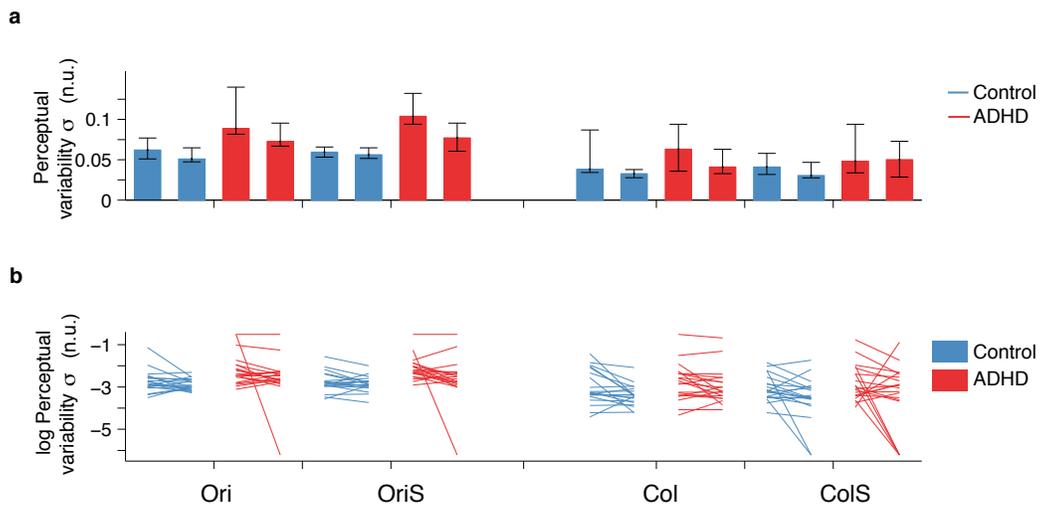
For the log lapse  $\lambda$ , we found a significant effect of feature ( $F(1, 38) = 25.88, p < 0.0001, \eta_p^2 = 0.40$ ) and a significant feature  $\times$  group interaction ( $F(1, 38) = 6.01, p = 0.02, \eta_p^2 = 0.14$ ); nothing else was significant ( $p > 0.09$ ). After Sidak correction ( $\alpha = 0.0043$ , all 12 comparisons make sense since  $\lambda$  is unitless), no between-group comparisons were significant ( $p > 0.02$ ). Within Controls, the feature comparisons Ori vs Col and OriS vs ColS were significant ( $p < 0.001$ ), but not the load ones. Within ADHD, neither the feature nor the load comparisons reached significance ( $p > 0.02$ ).

For the PSE  $\mu$ , like in the “shared” model, we found a significant effect of group ( $F(1, 38) = 10.85, p = 0.002, \eta_p^2 = 0.22$ ) and also a significant group  $\times$  load  $\times$  feature interaction ( $F(1, 38) = 8.42, p = 0.006, \eta_p^2 = 0.18$ ), nothing else reaching significance ( $p > 0.09$ ). After Sidak correction ( $\alpha = 0.0065$ , as for  $\sigma$ ), we found a significant difference between ADHD and Controls for ColS ( $p = 0.002$ ), but not for Col ( $p = 0.18$ ) and not for Ori or OriS ( $p > 0.12$ ). Again, these results cannot provide robust support for ADHD participants having different S-cone dependent color processing.



## Effect of learning

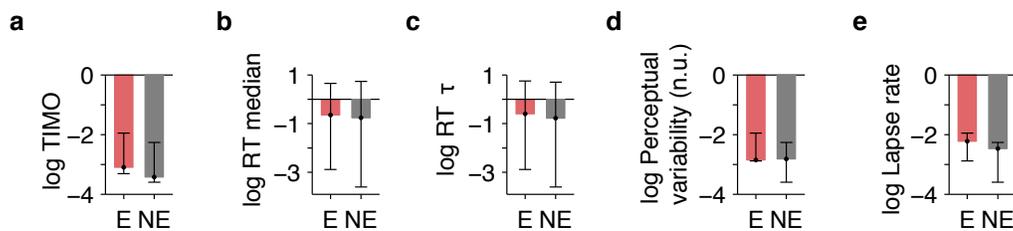
To assess learning across the experiment, we looked at the parameter estimates from the first half of the trials versus the second. Figure 3.14 shows that the perceptual variability parameters improved slightly on the second versus first half of trials, sign that there might be some learning. As reported in main, four-way mixed-design ANOVA on  $\log \sigma$  confirmed a significant effect of time:  $F(1, 38) = 12.7, p = 0.001, \eta_p^2 = 0.25$ . However, we note that these parameter estimates are not as reliable as the ones in Figure 3.4B, since they were obtained by fitting on only half the data.



**Figure 3.14** Perceptual variability parameter fits across time. **(a)** Medians across participants and bootstrapped 95% confidence intervals. **(b)** log Perceptual variability values of individual participants. Lines show how the perceptual variability differs from the first half of trials to the second half.

### 3.5.5 Effect of eye tracking

A possible concern is that half of the participants in each group were eye-tracked, while half were not. If an eye-tracked participant broke fixation, they had to redo the trial. As a result, the eye-tracked participants started more trials (mean and SD for eye-tracked:  $1047 \pm 201$  trials; non-eye-tracked always completed 800 trials). Thus, a concern could be that differences in task metrics could simply arise due to the experiment being longer and as a result more tiresome. We examined each of the average task metrics within each group, separately for the eye-tracked participants and the non-eye-tracked ones and found no significant differences (Wilcoxon rank-sum  $p > 0.13$ ) (Figure 3.15).



**Figure 3.15** No significant difference between eye tracked (E) and non-eye tracked (NE) participants on behavioral task metrics: (a) TIMO, (b) RT median, (c) RT  $\tau$ , (d) Perceptual variability, (e) Lapse rate. Bars represent medians and error bars bootstrapped 95% confidence intervals.

### 3.5.6 Breakdown of correlations

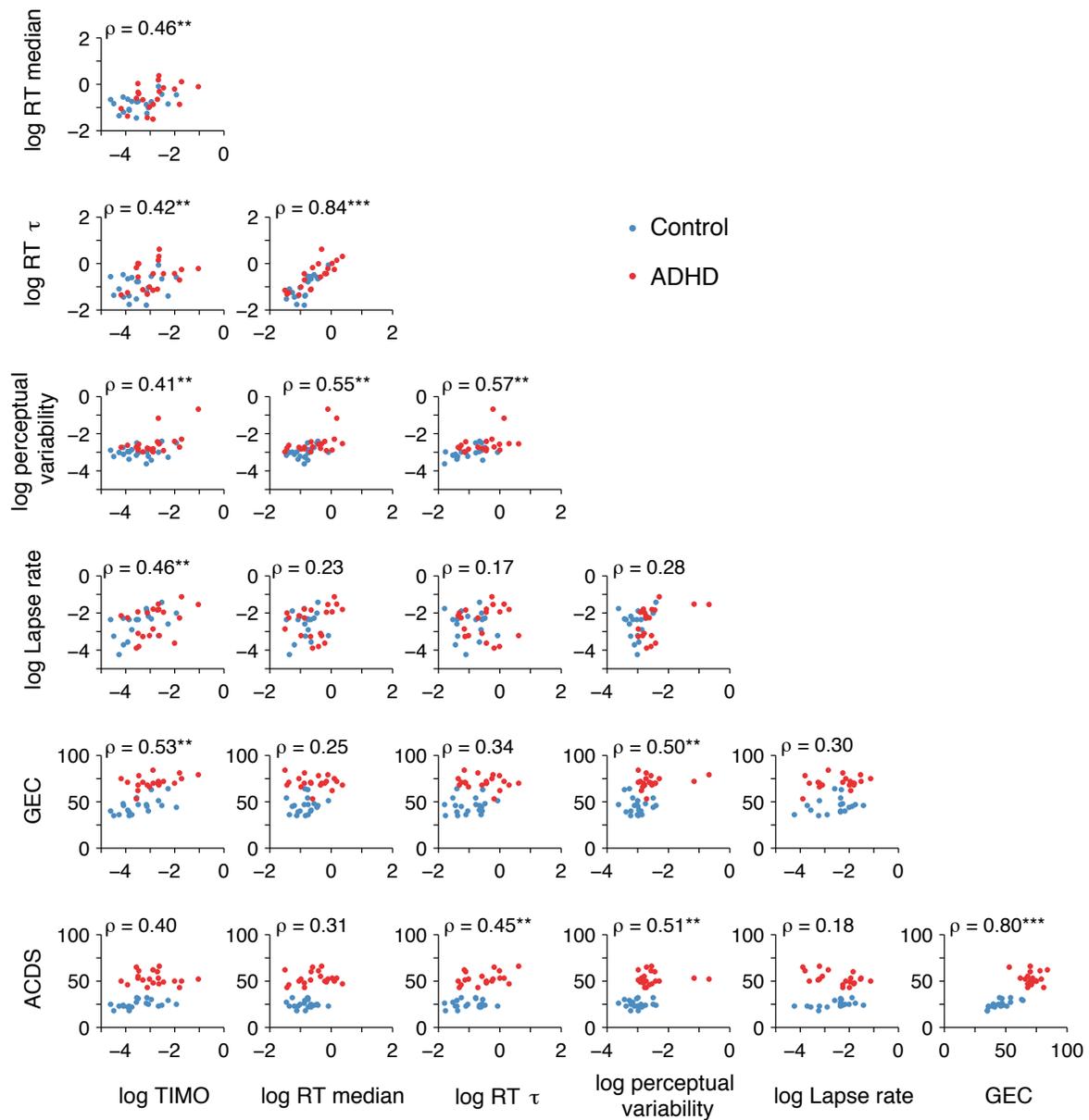
#### By group

In Figure 3.16, we show the points that make up the correlations from Table 3.1, color coded by group. Of note, the two ADHD participants who had visibly lower orientation discrimination performance (Figure 3.4A), did not also have outstandingly reduced perfor-

mance on other metrics; more detailed ophtamological examination could have provided more insight into the possible sources of their reduced orientation discrimination performance.

In Table 3.5 we show the pairwise correlations across task metrics separately within the Control group and within the ADHD group. Here, the only group specific correlations that survive the multiple-comparisons correction are RT with RT  $\tau$  and GEC with ACDS.

In addition, we attempted to determine whether for a given pair of task metrics, their correlation within the ADHD group is different from their correlation within the Control group. To do this, we compared the difference between the actual correlations to a distribution of differences between correlations obtained by shuffling the ADHD and Control labels. We did not find significant differences between the ADHD and Control correlations for any pair of task metrics ( $p > 0.04$ ).



**Figure 3.16 Dots: pairwise task metrics, color coded by group.** We also show here the Spearman correlations collapsed across groups, as in Table 3.1. **\*\*** depicts  $0.001 < p < 0.0089$  (since  $\alpha_{\text{Sidak}} = 0.0089$  after multiple-comparisons correction) and **\*\*\*** depicts  $p < 0.001$ .

**Table 3.5 No evident pattern of group specific correlations.** Pairwise Spearman correlations across task metrics (both behavioral and clinical), as in Table 3.1, but divided by group. Boldfaced is significant after multiple-comparisons correction,  $\alpha = 0.0083$  for Control and  $\alpha = 0.0082$  for ADHD. (see Methods)

(a) Control

	TIMO	RT	RT $\tau$	Perceptual variability ( $\sigma$ )	Lapse ( $\lambda$ )	GEC
TIMO						
RT	$\rho = 0.25$ $p = 0.27$					
RT $\tau$	$\rho = 0.13$ $p = 0.59$	<b><math>\rho = 0.81</math></b> $p < 0.0001$				
Perceptual variability ( $\sigma$ )	$\rho = -0.02$ $p = 0.92$	$\rho = 0.52$ $p = 0.02$	$\rho = 0.50$ $p = 0.03$			
Lapse rate ( $\lambda$ )	$\rho = 0.49$ $p = 0.03$	$\rho = 0.22$ $p = 0.34$	$\rho = 0.19$ $p = 0.42$	$\rho = -0.05$ $p = 0.84$		
GEC	$\rho = 0.52$ $p = 0.02$	$\rho = 0.16$ $p = 0.49$	$\rho = 0.15$ $p = 0.52$	$\rho = -0.12$ $p = 0.61$	$\rho = 0.25$ $p = 0.29$	
ACDS	$\rho = 0.49$ , $p = 0.03$	$\rho = -0.09$ $p = 0.67$	$\rho = 0.03$ $p = 0.89$	$\rho = -0.09$ $p = 0.69$	$\rho = 0.46$ $p = 0.04$	<b><math>\rho = 0.70</math></b> $p < 0.0001$

(b) ADHD

	TIMO	RT	RT $\tau$	Perceptual variability ( $\sigma$ )	Lapse ( $\lambda$ )	GEC
TIMO						
RT	$\rho = 0.48$ $p = 0.03$					
RT $\tau$	$\rho = 0.36$ $p = 0.12$	<b><math>\rho = 0.82</math></b> $p < 0.0001$				
Perceptual variability ( $\sigma$ )	$\rho = 0.46$ $p = 0.04$	$\rho = 0.52$ $p = 0.02$	$\rho = 0.46$ $p = 0.04$			
Lapse rate ( $\lambda$ )	$\rho = 0.40$ $p = 0.08$	$\rho = 0.38$ $p = 0.10$	$\rho = 0.12$ $p = 0.61$	$\rho = 0.32$ $p = 0.16$		
GEC	$\rho = 0.37$ $p = 0.11$	$\rho = -0.10$ $p = 0.68$	$\rho = 0.19$ $p = 0.42$	$r = 0.15$ $p = 0.54$	$r = 0.18$ $p = 0.44$	
ACDS	$\rho = -0.18$ $p = 0.45$	$\rho = 0.14$ $p = 0.56$	$\rho = 0.36$ $p = 0.11$	$\rho = 0.03$ $p = 0.87$	$\rho = -0.35$ $p = 0.12$	$\rho = -0.09$ $p = 0.07$

By symptom type

For this analysis, 2 ADHD participants were excluded due to missing AISRS records. A breakdown of the AISRS scores into inattentive and hyperactive shows that their correlations with task metrics recapitulate the correlations seen with ACDS. This is not unexpected, given the high correlation between ACDS and AISRS scores, as well as the

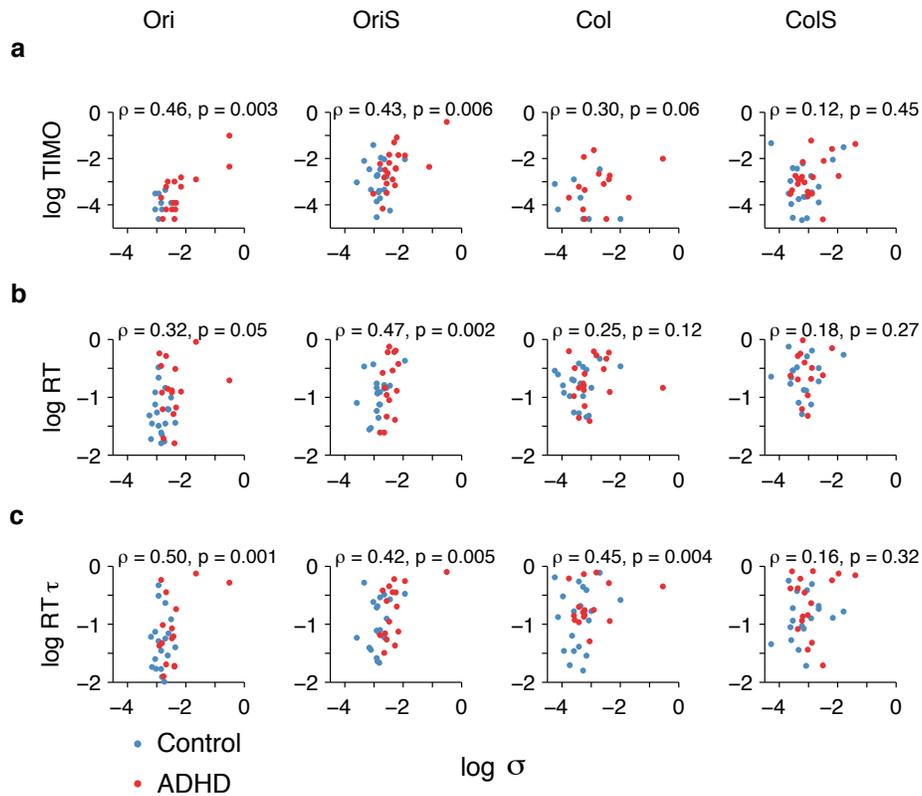
fact that the AISRS inattentive and AISRS hyperactive scores were highly correlated ( $r = 0.89, p < 10^{-13}$ ).

**Table 3.6 No evident pattern of differential correlations by symptom type.** AISRS inattentive and hyperactive correlations with behavioral task metrics are almost identical and largely recapitulate the ACDS correlations. Boldfaced represents  $p < 0.0089$ .

	TIMO	RT	RT $\tau$	Perceptual variability ( $\sigma$ )	Lapse ( $\lambda$ )	GEC
ACDS	$\rho = 0.40$ $p = 0.01$	$\rho = 0.31$ $p = 0.05$	<b><math>\rho = 0.45</math></b> $p = 0.004$	<b><math>\rho = 0.51</math></b> $p = 0.0008$	$\rho = 0.18$ $p = 0.26$	<b><math>\rho = 0.80</math></b> $p < 0.0001$
AISRS inattentive	$\rho = 0.41$ $p = 0.01$	<b><math>\rho = 0.43</math></b> $p = 0.006$	<b><math>\rho = 0.46</math></b> $p = 0.003$	<b><math>\rho = 0.51</math></b> $p = 0.001$	$\rho = 0.22$ $p = 0.17$	<b><math>\rho = 0.65</math></b> $p < 0.0001$
AISRS hyperactive	$\rho = 0.41$ $p = 0.01$	<b><math>\rho = 0.48</math></b> $p = 0.002$	<b><math>\rho = 0.49</math></b> $p = 0.001$	<b><math>\rho = 0.63</math></b> $p < 0.0001$	$\rho = 0.18$ $p = 0.27$	<b><math>\rho = 0.65</math></b> $p < 0.0001$

### By condition

In Table 3.1, for each participant, we averaged each behavioral metric across all four conditions. In Figure 3.17, we present the correlations of perceptual variability with TIMO, RT and RT  $\tau$  broken down by condition. Overall, we cannot conclude much from these patterns of results.



**Figure 3.17** Spearman correlations of perceptual variability with other behavioral metrics broken down by conditions show no conclusive pattern. We show the correlations of log perceptual variability with (a) log TIMO, (b) log RT median and (c) log RT  $\tau$ .

### 3.5.7 Prediction of clinical variables

Logistic regression: prediction of diagnosis from behavioral metrics

**Table 3.7** Logistic regression coefficients, mean  $\pm$  sem.

**(a) Diagnosis  $\sim$  log perceptual variability**

	coefficient	<i>t</i> value	<i>p</i> value
intercept	13.3 $\pm$ 4.8	2.78	0.0055**
log perceptual variability ( $\sigma$ )	4.7 $\pm$ 1.7	2.80	0.0051**

**(b) Diagnosis  $\sim$  log TIMO**

	coefficient	<i>t</i> value	<i>p</i> value
intercept	1.9 $\pm$ 0.9	2.07	0.038*
log TIMO	1.1 $\pm$ 0.5	2.25	0.024*

**(c) Diagnosis  $\sim$  log perceptual variability + log TIMO**

	coefficient	<i>t</i> value	<i>p</i> value
intercept	13.3 $\pm$ 4.9	2.70	0.0069**
log perceptual variability ( $\sigma$ )	4.3 $\pm$ 1.7	2.47	0.013*
log TIMO	0.63 $\pm$ 0.59	1.07	0.28

**(d) Diagnosis  $\sim$  log perceptual variability + log TIMO + log RT median + log RT  $\tau$  + log lapse rate**

	coefficient	<i>t</i> value	<i>p</i> value
intercept	14.0 $\pm$ 5.4	2.57	0.010*
log TIMO	0.49 $\pm$ 0.64	0.76	0.44
log RT median	-1.4 $\pm$ 1.7	-0.81	0.42
log RT $\tau$	1.4 $\pm$ 1.6	0.88	0.38
log perceptual variability ( $\sigma$ )	4.2 $\pm$ 1.9	2.23	0.025*
log lapse rate ( $\lambda$ )	0.38 $\pm$ 0.37	1.02	0.30

Linear regression: prediction of clinical metrics GEC and ACDS from behavioral metrics

**Table 3.8** Linear regression coefficients, depicted as mean  $\pm$  sem, for GEC and ACDS with task metrics.

(a) GEC $\sim$ log perceptual variability + log TIMO + log RT + log RT $\tau$ + log lapse rate.				(b) ACDS $\sim$ log perceptual variability + log TIMO + log RT + log RT $\tau$ + log lapse rate.			
	coefficient	<i>t</i> value	<i>p</i> value		<i>t</i> value	<i>p</i> value	
log intercept	103 $\pm$ 12	8.50	$< 10^{-9***}$	73 $\pm$ 13	5.77	$< 10^{-5***}$	
log TIMO	8.0 $\pm$ 3.2	2.53	0.016*	3.4 $\pm$ 3.3	1.04	0.31	
log RT median	-10.3 $\pm$ 7.6	-1.35	0.18	-11.3 $\pm$ 8.0	-1.41	0.15	
log RT $\tau$	8.1 $\pm$ 6.1	1.32	0.19	14.4 $\pm$ 6.4	2.26	0.031*	
log $\sigma$	6.1 $\pm$ 4.9	1.23	0.22	6.8 $\pm$ 5.1	1.33	0.19	
log $\lambda$	1.3 $\pm$ 1.8	0.75	0.46	0.5 $\pm$ 1.9	0.29	0.78	

# Chapter 4

## Bayesian microsaccade detection with an application to ADHD

*The pure fire within us is akin to this, and they (the gods) caused it to flow through the eyes. Whenever there is daylight round about, the visual current issues forth, like to like, and coalesces with the daylight and is formed into a single homogenous body in a direct line with the eyes, in whatever quarter the stream issuing from within strikes upon any object it encounters outside.*

Plato, Timaeus, 360 B.C.

---

### 4.1 Introduction

In Chapter 2 and Chapter 3, we addressed aspects of covert attention through behavioral measures. Here, we examine a possible physiological correlate of covert attention, namely microsaccades. People's eyes are always in motion, even when they attempt to fix-

ate. These eye movements are not just random, but can be subject to behavioral control, serving to optimize the processing of visual information (Rucci and Victor, 2015). Thus, while Plato's extramission theory referred to in the quote above was regarded as wrong even by his contemporaries, in some metaphorically sense it can be read as an early allusion to the importance of oculomotor control for visual perception.

Fixational eye movements are often categorized as drift, tremor, and microsaccades (Ciuffreda and Tannen, 1995). Most of the fixation time, the eye is in a drift state, which is low-amplitude, low-velocity motion, sometimes modeled as a random walk (Cornsweet, 1956; Ditchburn and Ginsborg, 1953; Ratliff and Riggs, 1950). A few times a second, the eye makes a microsaccade, which is a high-velocity, high-amplitude movement along a relatively straight trajectory (Barlow, 1952; Cornsweet, 1956; Engbert and Kliegl, 2003; Engbert et al., 2011; Rolfs, 2009).

Microsaccades have been investigated and found to be implicated in several perceptual and cognitive functions, including aiding performance in high-acuity visual tasks (Ko et al., 2010; Poletti et al., 2013; Rucci et al., 2007) and shifts of covert spatial attention in both humans and monkeys (Engbert and Kliegl, 2003; Hafed and Clark, 2002; Hafed et al., 2011; Lara and Wallis, 2012; Laubrock et al., 2005, 2007; Rolfs, 2009; Rolfs et al., 2004; Yuval-Greenberg et al., 2014). For more details on these general roles, see Section 1.6.

#### 4.1.1 Microsaccades in ADHD

If microsaccades have special roles in perception and can be considered indices of covert attention, perhaps ADHD participants show differential patterns of microsaccades?

Only a few studies examined microsaccades in ADHD. One line of work proposed that superior colliculus function could be affected in ADHD (Overton, 2008), based on its role in orienting of attention and the generation of saccades, and ADHD participants' differential saccade patterns (Munoz et al., 2003). Indeed, a study found that participants with high ADHD traits performed more microsaccades during a sustained fixation task, and the microsaccade rate was correlated with symptom severity (Panagiotidi et al., 2017). In another line of work (Fried et al., 2014) recorded fixational eye movements during the widely used Continuous performance test (CPT) and found increased microsaccade and blink rates around the stimulus onset in the ADHD group, but not in the ADHD medicated group.

Beyond microsaccadic inhibition post stimulus onset, (Betta and Turatto, 2006) and others found that the expectation and preparation for a stimulus can decrease microsaccade rate even before the stimulus is presented. To address whether usage of temporal expectations is different in ADHD, (Dankner et al., 2017) made use of a variant of the CPT task with two conditions, fixed and variable interstimulus intervals (ISIs), in which the latter taxed sustained attention more. In contrast to (Fried et al., 2014), they found that neurotypicals made more microsaccades in the -100 ms to 0 ms prestimulus interval, difference possibly explained by using a central stimulus, while (Fried et al., 2014) used a peripheral (9 dva) stimulus. However, (Dankner et al., 2017) also found, as they expected, that the reliable temporal expectation in the fixed ISI condition gave rise to higher prestimulus saccadic inhibition (PSSI) than in the variable ISI condition. They defined the difference of PSSI from variable to fixed as the PSSI-predictability effect and saw that ADHD participants, and particularly participants with lower sustained attention, had lower PSSI-predictability effects. (Roberts et al., 2017) examined spatial covert attention, both endogenous and exogenous, with a 2AFC orientation discrimination task and found

no significant differences in ADHD participants relative to Controls neither in their attentional allocation abilities nor in their overall microsaccade rate. Future studies are needed to complement these results and construct a more detailed understanding of which tasks and parameters and which features of microsaccades are different in ADHD.

#### 4.1.2 Microsaccade detection methods

Arguments about the functional roles of microsaccades rely on the accurate definition and detection of microsaccades (Poletti and Rucci, 2016). Microsaccade detection is complicated by motor noise in the eye and by measurement noise in the eye tracker. The latter is particularly important in view of the widespread use of video-based infrared eye trackers, which are less invasive than magnetic scleral search coils, but noisier (Hermens, 2015; Träisk et al., 2005). For example, the popular EyeLink II video-based infrared eye tracker reports a precision of 0.01 degrees of visual angle; however, in practice this precision can be worse (Holmqvist et al., 2011). The low sensitivity, precision and resolution of video-based eye trackers can cause difficulties in resolving microsaccades (Nyström et al., 2016; Poletti and Rucci, 2016).

How can microsaccades be reliably detected in the presence of other fixational eye movements and measurement noise? The most commonly used microsaccade detection method, especially in human studies, is a velocity threshold algorithm proposed by Engbert and Kliegl (Engbert, 2006; Engbert and Kliegl, 2003). This method, which we refer to as EK, detects a microsaccade when the magnitude of the eye velocity exceeds a given threshold for a sufficiently long duration. Because a fixed threshold would ignore differences in noise across trials and individuals, the authors adaptively choose the threshold to be a multiple of the standard deviation of the velocity distribution (Engbert and Kliegl,

2003). However, the value of the multiplier is arbitrary and affects the algorithm's performance, as expected from signal detection theory: If the multiplier is too high, the algorithm misses microsaccades, while too low a multiplier causes false alarms. For example, EK with the threshold multiplier set to its standard value of 6 (Engbert and Kliegl, 2003) labels the eye position data in Figure 4.1A as a microsaccade, but not the data in Figure 4.1B. However, lowering the threshold multiplier to 3 causes EK to label both examples as microsaccades. This ambiguity in the identification of microsaccades can cause ambiguity in conclusions about their functional roles.

To increase the confidence that a fixational eye movement sequence is indeed a microsaccade and reduce the false positives induced by threshold reduction, it has become a common approach in the case of binocular datasets to detect microsaccades independently in each eye and reject as noise microsaccades that are detected only in one eye (Engbert, 2006; Engbert and Mergenthaler, 2006; Otero-Millan et al., 2008). That monocular microsaccades are indeed mostly safe to exclude has been confirmed more recently by (Fang et al., 2018; Nystrom et al., 2017).

More recent algorithms have tried to eliminate the need for an arbitrary velocity threshold by taking into account more details of the statistics of fixational eye movements. (Bettenbuehl et al., 2010) assume that microsaccades are discontinuities embedded in drift and use wavelet analysis to detect them. (Otero-Millan et al., 2014a) proposed an unsupervised clustering algorithm based on three features: peak velocity, initial acceleration peak and final acceleration peak. This clustering method substantially reduced detection errors for binocular data as well as for monocular data.

### 4.1.3 Bayesian methods for saccade and microsaccade detection

Bayesian algorithms have been used previously specifically for saccade detection. (Salvucci and Anderson, 1998; Salvucci and Goldberg, 2000) used a hidden Markov model to separate fixations from saccades. However, their algorithm requires the user to specify a set of fixation targets, which are regions of interest based on a cognitive process model of the task. By contrast, our algorithm is entirely task-independent. More recently, (Daye and Optican, 2014) used particle filters to estimate the posterior over a hidden position variable in a generic and simple model for eye velocity. Whenever this model fails to capture the data, their algorithm concludes that a microsaccade or saccade has occurred. Instead, we build an explicit model of both microsaccades and drift, and compute the full posterior over the eye state. (Santini et al., 2016) proposed to use a Bayesian classifier to separate fixations, saccades, and smooth pursuits based on two features: eye speed and a feature which distinguishes smooth from abrupt motion. This algorithm was applied to much lower resolution eye tracking data (30 Hz) than typically used in psychology laboratories and while principled, still relied on pre-processing using a heuristic filter. This method seems to work when separating saccades from drift, but we focus on the harder problem of separating microsaccades from drift.

### 4.1.4 Our approach

We started out trying to apply the Bayesian online changepoint detection algorithm by Adams and MacKay (Adams and McKay, 2007) to identify microsaccades. Change-points partition data time series into sequences that can be captured with specific parameters of Gaussian distributions, just as when the eye starts a microsaccade sequence the

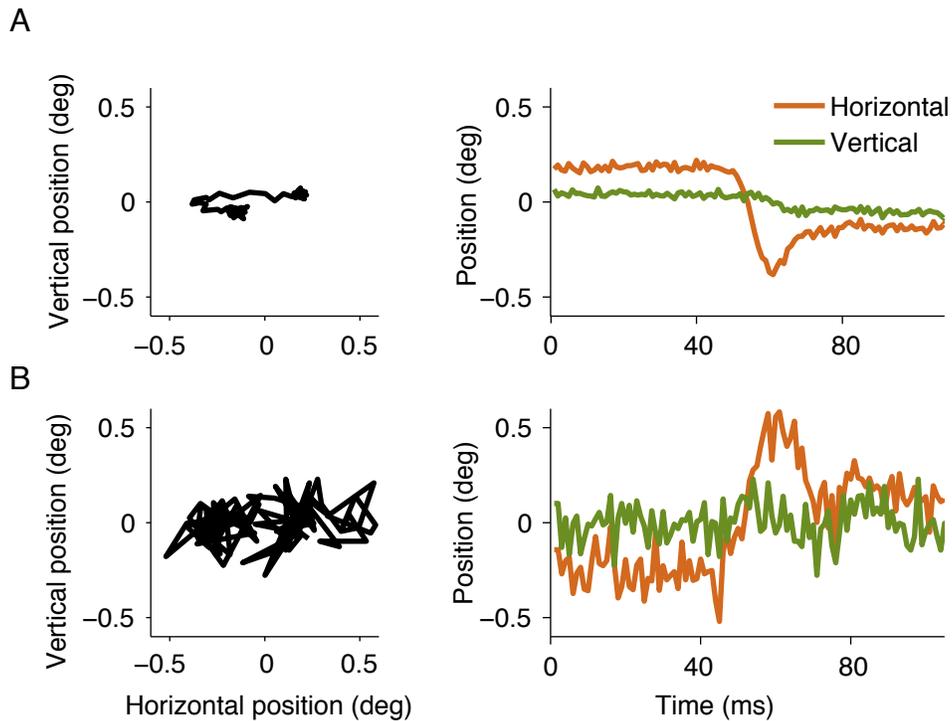
statistics of the velocity distribution changes. Due to the short duration of microsaccades, such an online algorithm might not have enough samples to confidently infer that the eye has entered a new velocity distribution with new parameters.

Fortunately, for many applications, it is not necessary to detect microsaccades online. Making use of this fact, we have developed a new method, Bayesian microsaccade detection (BMD), which performs inference post hoc based on a simple statistical model of eye positions. In this model, a hidden state variable changes between drift and microsaccade states at random times. The eye position is a biased random walk with different velocity distributions for each state. BMD generates samples from the posterior probability distribution over the eye state time series given the eye position time series.

#### **4.1.5 Structure of this chapter**

Section 4.2.1, Section 4.2.2 and Section 4.2.3 consist of an explanation of the formalism of BMD. Section 4.2.4 presents the alternative microsaccade detection algorithms, and the rest of the Methods section contain data analysis and data collection methods. In Section Section 4.3.1, we apply BMD to simulated data and compare the performance to alternative algorithms Section 4.2.4 at varying levels of noise. In Section 4.3.2, we apply BMD to pilot EyeLink eye tracker data, which have relatively high noise. In Section 4.3.3, we apply BMD to Dual Purkinje Image eye tracker, whose higher precision we assume justifies defining the inferred microsaccades from all the algorithms as ground truth. We present variants of BMD in Section 4.3.4. In Section 4.3.5, we apply BMD to the EyeLink dataset collected while ADHD and control participants performed the task described in Chapter 3. In Section 4.4.1, we discuss the limitations of BMD. Lastly, in Section 4.4.3, we comment on the applicability of connecting microsaccades with behavioral measures in

the study of ADHD.



**Figure 4.1** Microsaccades under different noise levels. Example single-trial eye position data from two subjects, measured with the EyeLink eye tracker with the “Heuristic filter” option turned off. **(A)** Measured eye position in the plane (left) and horizontal and vertical position as a function of time (right) for an easily detectable microsaccade. **(B)** Another trace, which contains an apparent microsaccade buried in measurement noise. EK with the threshold multiplier set at 6 identifies a microsaccade in **A**, but not in **B**.

## 4.2 Methods

We develop a Bayesian algorithm for microsaccade detection. First, we make explicit assumptions about the statistical process by which the eye movement data are generated from an underlying sequence of hidden states alternating between drift and microsaccades. Optimal Bayesian inference then entails inverting this generative model to infer the proba-

bility of the hidden eye state sequence given the measured eye position data. The fact that our algorithm returns a probability distinguishes it from earlier algorithms, which return only a binary judgment.

The input of our inference algorithm is a time series of measured eye positions. We conceptualize this time series as being generated from an unknown internal state, which at each time step is either drift/tremor (0) or microsaccade (1). We distinguish the two internal states by asserting that they correspond to different velocity distributions; this statistical definition stands in contrast to the traditional method, which uses a threshold. The probability distributions that describe the process by which the measured eye position time series arises from the internal state are together called the generative model.

Assuming this generative model, we derive an inference algorithm that estimates the time series of hidden eye states given a particular measured eye position time series. The algorithm considers many candidate time series (e.g. 0001111100...00111111111000) and calculates how consistent each candidate is with the data; this is called the likelihood of that candidate time series. Combining the likelihoods with prior information about frequencies and durations of microsaccades yields the posterior distribution over time series. Because the space of candidate time series is very large  $-2^{60000}$  for 1 minute of data sampled at 1 kHz, we use a suitable search algorithm from the class of Markov-Chain Monte Carlo (MCMC) sampling algorithms. A computer package implementing our algorithm is available at

<https://github.com/basvanopheusden/BMD>

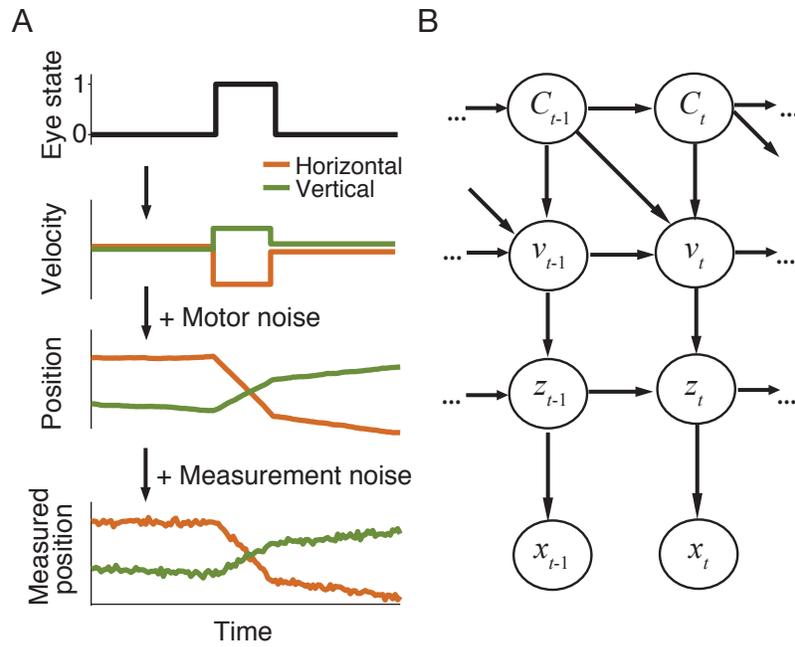
### 4.2.1 Generative model

We formulate our model to generate eye position data in 1 ms time bins, but this can be easily extended to different sampling rates. We use boldface to denote a time series of a variable. The eye state time series  $\mathbf{C}$  has length  $T$ . We assume that at time  $t$ , the eye is in a hidden state  $C_t$ , which is 0 for a “drift/tremor” state and 1 for a “microsaccade” state (Figure 4.2). As long as the eye remains in the same state, its two-dimensional velocity  $v_t$  remains constant; when the eye switches state, its velocity changes. Of note, velocity here does not represent the derivative of the measured eye position, but an unobservable underlying variable.

At every time step, the eye’s two-dimensional position  $z_t$  gets augmented with the velocity and Gaussian motor noise with covariance matrix  $\Sigma_z$ . This eye position is augmented with Gaussian measurement noise with covariance matrix  $\Sigma_x$  (independent across time points), yielding the measured eye position  $x_t$ .

We define a change point of  $\mathbf{C}$  as a time  $t$  where  $C_t \neq C_{t-1}$ , and denote the  $i^{\text{th}}$  change point by  $\tau_i$ . The duration for which the eye stays in a given state is then  $\Delta\tau_i = \tau_{i+1} - \tau_i$ . We assume that  $\mathbf{C}$  is a semi-Markov process, which means that these durations are independent. In a hidden Markov model (Bishop, 2006), the probability of  $C_t$  only depends on the previous state,  $C_{t-1}$ ; however, in a hidden semi-Markov model (also called explicit-duration hidden Markov model) (Yu, 2010), the durations over which the state remains unchanged are independent. Then, the prior probability of  $\mathbf{C}$  is

$$p(\mathbf{C}) = \prod_i p(\Delta\tau_i | C_{\tau_i}), \tag{4.1}$$

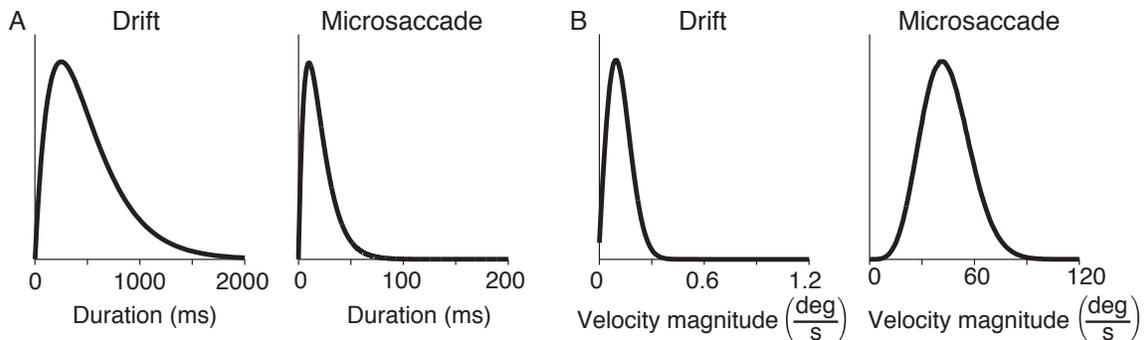


**Figure 4.2** Generative model of fixational eye movements. **(A)** Example time courses of the variables in our model. **(B)** Graphical representation of our generative model for eye position data, which is a hidden semi-Markov model. The eye state, a latent binary variable  $C_t$ , is either a low-velocity ‘drift/tremor’ state (0) or a high-velocity ‘microsaccade’ state (1). The latent eye state informs the velocity  $v_t$ , which together with the preceding eye position  $z_{t-1}$  and motor noise yield the current eye position  $z_t$ . This eye position is contaminated with measurement noise, yielding the measured eye position  $x_t$ .

where  $p(\Delta\tau_i|C_{\tau_i})$  is the state-specific probability of the duration. Specifically, we use a gamma distribution with shape parameter 2 and scale parameter  $k$ :

$$p(\Delta\tau|C) \propto \Delta\tau e^{-k\Delta\tau}, \quad (4.2)$$

where  $k = k_0$  if  $C = 0$  (drift/tremor) and  $k = k_1$  if  $C = 1$  (microsaccade). We choose this distribution because it makes very short and very long durations unlikely, consistent with previously reported distributions of durations for drift and microsaccades (Engbert, 2006). Assumptions about the frequency and duration of microsaccades are reflected in the choices of parameters  $k_0$  and  $k_1$ . We chose  $k_0 = 4 \text{ s}^{-1}$  and  $k_1 = 100 \text{ s}^{-1}$ , corresponding to median durations of 260 ms for drift and 10 ms for microsaccades (Figure 4.3A,B), which are realistic (Engbert, 2006). We will later examine the robustness of our results to variations in  $k_0$  and  $k_1$  (Figure Figure 4.22).



**Figure 4.3 Prior distributions used in the algorithm.** (A) Prior distributions over the durations of drift and microsaccade states. These priors are fixed in the inference process. (B) Priors for eye velocity for drift and microsaccade states. The inference process estimates the parameters of these priors from data; here we show the priors estimated for one example subject (EyeLink S1, Table 4.2). Note that these distributions are not normalized.

At each change point  $\tau_i$ , we draw the velocity  $v_{\tau_i}$  from a state-specific probability distri-

bution  $p(v_{\tau_i}|C_{\tau_i})$ ; this velocity remains constant until the eye switches state at  $\tau_{i+1}$ . The distribution of the velocity time series  $\mathbf{v}$  given an eye state time series  $\mathbf{C}$  is

$$p(\mathbf{v}|\mathbf{C}) = \prod_i \left[ p(v_{\tau_i}|C_{\tau_i}) \prod_{t=\tau_i+1}^{\tau_{i+1}-1} \delta(v_t - v_{t-1}) \right] \quad (4.3)$$

To define the state-specific velocity distribution, we write  $v$  in polar coordinates,  $v = (r \cos \theta, r \sin \theta)^\top$ , and assume that in both states, the direction of the velocity  $\theta$  is uniformly distributed, and its magnitude  $r$  follows a generalized gamma distribution:

$$p(r|C) = \frac{2}{\left( \Gamma\left(\frac{d+1}{2}\right) (\sigma\sqrt{2})^{d+1} \right)} r^d e^{-\frac{r^2}{2\sigma^2}} \quad (4.4)$$

where  $d = d_0$  and  $\sigma = \sigma_0$  if  $C = 0$ , and  $d = d_1$  and  $\sigma = \sigma_1$  if  $C = 1$ . Note that our definition of the generalized gamma distribution differs from that in (Stacy, 1962) by a reparametrization  $d \rightarrow d + 1, \sigma \rightarrow \sigma\sqrt{2}$ . We fix  $d_0$  to 1, which is equivalent to assuming that the distribution of the two-dimensional velocity in the drift/tremor state is a circularly symmetric Gaussian with standard deviation  $\sigma_0$ . The other parameters  $d_1$  and  $\sigma_1$  control the shape and scale of the distribution of microsaccade velocities, respectively. Figure 4.3B shows examples of these velocity distributions.

The eye position time series  $\mathbf{z}$  is piecewise linear with velocity  $\mathbf{v}$ , plus motor noise, which follows a Gaussian distribution with covariance matrix  $\Sigma_z$ :

$$p(\mathbf{z}|\mathbf{v}) = \prod_{t=1}^T p(z_t|z_{t-1}, v_t) = \prod_{t=1}^T \mathcal{N}(z_t; z_{t-1} + v_t, \Sigma_z) \quad (4.5)$$

The observed eye position time series  $\mathbf{x}$  is equal to  $\mathbf{z}$  plus Gaussian measurement noise that is independent across time and has covariance matrix  $\Sigma_x$ :

$$p(\mathbf{x}|\mathbf{z}) = \prod_{t=1}^T p(x_t|z_t) = \prod_{t=1}^T \mathcal{N}(x_t; z_t, \Sigma_x) \quad (4.6)$$

Motor and measurement noise are in principle distinguishable, because changes in the eye position due to motor noise are added over time, whereas the measurement noise is independently added at each time point. We assume that both covariance matrices are isotropic:  $\Sigma_z = \sigma_z^2 \mathbf{I}$  and  $\Sigma_x = \sigma_x^2 \mathbf{I}$ . Before we analyze data, we rescale the vertical dimension of the measured eye positions so that the isotropy assumption is approximately satisfied (see Preprocessing).

#### 4.2.2 Inference of the eye state time series

Our goal is to infer the eye state time series  $\mathbf{C}$  given a time series of measured eye positions  $\mathbf{x}$ , using the generative model above. To perform optimal inference, we need to compute the posterior distribution over  $\mathbf{C}$ . By Bayes' rule, this posterior is proportional to the product of the prior  $p(\mathbf{C})$  and the likelihood  $p(\mathbf{x}|\mathbf{C})$ :

$$p(\mathbf{C}|\mathbf{x}) \propto p(\mathbf{C})p(\mathbf{x}|\mathbf{C}).$$

The prior can be directly evaluated using Equation (4.1) and Equation (4.2), but computing the likelihood requires marginalization over nuisance parameters, the velocity time series  $\mathbf{v}$  and the eye position time series  $\mathbf{z}$ , using the dependencies given by the generative model.

$$p(\mathbf{x}|\mathbf{C}) = \iint p(\mathbf{x}|\mathbf{z})p(\mathbf{z}|\mathbf{C}, \mathbf{v})p(\mathbf{v}|\mathbf{C})d\mathbf{v}d\mathbf{z} \quad (4.7)$$

Plugging in the functional form of these distributions, and performing some algebra (see Supplementary. 4.5.1), yields the likelihood of  $\mathbf{C}$ :

$$\begin{aligned}
p(\mathbf{x}|\mathbf{C}) &\propto \int \prod_t \left[ e^{-\frac{(z_t - z_{t-1})^\top (z_t - z_{t-1})}{2\sigma_z^2} - \frac{(x_t - z_t)^\top (x_t - z_t)}{2\sigma_x^2}} \right] \\
&\prod_i \left[ \int p(v_{\tau_i} | C_{\tau_i}) e^{-\frac{(\tau_{i+1} - \tau_i) v_{\tau_i}^\top v_{\tau_i}}{2\sigma_z^2} + \frac{(z_{\tau_{i+1}} - z_{\tau_i})^\top v_{\tau_i}}{\sigma_z^2}} dv_{\tau_i} \right] d\mathbf{z} \quad (4.8)
\end{aligned}$$

### 4.2.3 Approximate inference

The goal of our algorithm is to draw samples from the posterior  $p(\mathbf{C}|\mathbf{x})$ . First, we need to evaluate the likelihood in Equation (4.8). This is difficult, because we need to integrate both over the velocities at all change points,  $\{v_{\tau_i}\}$ , and over the eye position time series,  $\mathbf{z}$ . The velocity integral is numerically tractable, but the eye position one is not. Moreover, the likelihood also depends on the unknown parameters  $\sigma_0, \sigma_1, d_1, \sigma_z$  and  $\sigma_x$ . A fully Bayesian algorithm would require priors over these parameters to jointly infer the parameters together with the eye state time series  $\mathbf{C}$ . This too is intractable.

Instead, we use a multi-step approximate inference algorithm, which we name Bayesian microsaccade detection (BMD), outlined in Table 4.1. A key idea in this algorithm is to replace the marginalization over  $\mathbf{z}$  by a single estimate, reminiscent of expectation maximization. Our algorithm then alternates between estimating  $\mathbf{C}$ ,  $\mathbf{z}$ , and the parameters for 6 iterations, which in practice suffices for the estimates to converge. To run BMD on an eye position time series of 1 minute (60,000 time steps) takes approximately 40 seconds on a MacBook Pro with an Intel Core i7 with a 2.9 GHz processor and 8 GB of RAM.

Although BMD returns a probability over eye state at every time point, for most of the

following analyses we will threshold these probabilities at 0.5 in order to obtain binary judgments.

Step	Operation
0	Initialize $\mathbf{C}, \sigma_1, d_1, \sigma_0$
1	Estimate the motor and measurement noise, $\hat{\sigma}_z, \hat{\sigma}_x$
2	Estimate $\hat{\mathbf{z}}$ from observations $\mathbf{x}$ : Kalman smoother
3	Sample from the posterior over $\mathbf{C}$ : MCMC
4	Estimate the velocity distribution parameters : MLE
	Return to Step 1

**Table 4.1 BMD algorithm.** Here and elsewhere, we denote estimates of variables with hats.

We now describe the details of the steps of the BMD algorithm.

#### 4.2.3.1 Preprocessing

We split the eye position data into  $\approx 1$  minute blocks, which we process independently. Before we perform inference, we preprocess the data to match the isotropy assumption of the measurement and motor noise in our generative model. To do so, we observe that within our model, eye velocity is piecewise constant, and therefore, its derivative is zero except at change points. This means that the acceleration of the measured eye position depends only on the motor and measurement noise, except at change points. For this reason, we use the median absolute deviation of the acceleration to estimate the noise level. We calculate this quantity separately in the horizontal and vertical dimensions, and rescale the vertical position time series by the ratio of the outcomes. After rescaling, the noise distribution is approximately isotropic.

The algorithm utilizes measured eye position at boundary unobserved time points  $x_0$  and  $x_{T+1}$ . For these, we choose  $x_0 = x_1 - \varepsilon$  and  $x_{T+1} = x_T + \varepsilon$ , where  $\varepsilon = (10^{-4}, 10^{-4})^T$  deg. We need to include the offset  $\varepsilon$  to avoid numerical instabilities in our implementation. Finally, we subtract the resulting value of  $x_0$  from every point in the time series, so that  $x_0 = 0$ ; this has no effect on the detected microsaccades.

#### 4.2.3.2 Step 0. Initialize $\mathbf{C}$ , $\sigma_1, d_1, \sigma_0$

We fix  $d_0$  to 1. We initialize  $\hat{\sigma}_0$ ,  $\hat{\sigma}_1$ , and  $\hat{d}_1$  to random values drawn from reasonable ranges ( $\hat{\sigma}_0$ :  $[0.0001, 0.005] \frac{\text{deg}}{\text{ms}}$ ,  $\hat{\sigma}_1$ :  $[0.005, 0.1] \frac{\text{deg}}{\text{ms}}$  and  $\hat{d}_1$ :  $[1.1, 5]$ ). We initialize  $\mathbf{C}$  by setting  $C_t$  to 1 for time points  $t$  where  $\|x_t - x_{t-1}\|$  is in the highest percentile, and to 0 otherwise.

#### 4.2.3.3 Step 1. Estimate the motor and measurement noise

Our first goal is to estimate  $\sigma_x$  and  $\sigma_z$  given a measured eye position time series  $\mathbf{x}$  and an estimated eye state time series  $\hat{\mathbf{C}}$ . As stated before, we can disentangle motor and measurement noise because, in our generative model, motor noise accumulates over time while measurement noise does not. Specifically, the autocovariance function of  $\mathbf{x}$  conditioned on  $\mathbf{v}$  at time lag  $s$  is

$$\text{cov}(x_t, x_{t-s}) = 2\sigma_z^2 s + 4\sigma_x^2.$$

To use this relationship, we first estimate  $\mathbf{v}$  by fitting  $\mathbf{x}$  as a piecewise linear function with discontinuities at the change points of  $\mathbf{C}$ . Then we calculate the empirical autocovariance function of the residual

$$c_{\text{emp}}(s) = \frac{1}{T} \sum_{t=s+1}^T \tilde{x}_t^T \tilde{x}_{t-s},$$

and fit this as a linear function of  $s$ ; this gives a slope and a y-intercept. Our estimates of the motor noise and measurement noise are  $\hat{\sigma}_z = \sqrt{\frac{\text{slope}}{2}}$  and  $\hat{\sigma}_x = \sqrt{\frac{\text{y-intercept}}{4}}$ .

#### 4.2.3.4 Step 2. Estimate $\mathbf{z}$ from observations $\mathbf{x}$ with Kalman smoother

We cannot compute the likelihood of the eye state time series,  $p(\mathbf{x}|\mathbf{C})$  in Equation (4.8), because the integral over  $\mathbf{z}$  is both analytically and numerically intractable. However, the integral over  $v_{\tau_i}$ , depends only on  $z_{\tau_{i+1}} - z_{\tau_i}$ . The expected value of this difference is equal to  $(\tau_{i+1} - \tau_i)\bar{v}$ , where  $\bar{v}$  is the average velocity between the change points, while its standard deviation is of the order of  $\sigma_z$ . Therefore, if either  $\bar{v}$  or  $\tau_{i+1} - \tau_i$  is sufficiently large (we expect the former to hold for microsaccades and the latter for drift), we can neglect the uncertainty in  $z_{\tau_{i+1}} - z_{\tau_i}$  and approximate it by a point estimate.

We obtain the point estimate of  $\mathbf{z}$  given  $\mathbf{x}$  by maximizing the first integral in Equation (4.8). This is suboptimal, but in practice it might turn out to be justified. This maximization turns out to be equivalent to applying a Kalman smoother to  $\mathbf{x}$  (Kalman, 1960; Welch and Bishop, 2006). In general, a Kalman smoother estimates the system state in a time interval from noisy observations during the same interval. The optimal estimate turns out to be a linear filter. We implement the Kalman smoother with the Rauch-Tung-Striebel (RTS) algorithm, which first applies a Kalman filter to  $\mathbf{x}$ , followed by another Kalman filter to the output of the first filter, backward in time (Rauch et al., 1965; Terejanu, 2008). The Kalman filter estimates the system state at each time only from earlier observations. In our case, the RTS algorithm reduces to

$$\hat{y}_t = \hat{y}_{t-1} + K_{\text{forward}}(x_t - \hat{y}_{t-1}), \quad (4.9)$$

with  $K_{\text{forward}} = \frac{1+\sqrt{1+4R^2}}{1+2R^2+\sqrt{1+4R^2}}$ , where  $R = \frac{\sigma_x}{\sigma_z}$ , and

$$\hat{z}_t = \hat{z}_{t+1} + K_{\text{backward}}(y_t - \hat{z}_{t+1}), \quad (4.10)$$

with  $K_{\text{backward}} = \frac{\sqrt{1+4R^2}-1}{\sqrt{1+4R^2}+1}$ . For more details, see Supplementary. 4.5.1.

Given our generative model, the Kalman smoother is the optimal filter to denoise the measured eye position. The EyeLink eye tracker software also has a denoising option, called ‘‘Heuristic filter’’, which is based on an algorithm by Stampe (Stampe, 1993a). This filter is suboptimal given our generative model and therefore, assuming that our generative model is realistic, it will perform worse in separating signal from noise than the Kalman smoother.

#### 4.2.3.5 Step 3. Sample from the posterior over the eye state time series $\mathbf{C}$

We draw samples from the posterior  $p(\mathbf{C}|\hat{\mathbf{z}}, \sigma_0, \sigma_1, d_1, \sigma_z)$  using Markov Chain Monte Carlo sampling with Metropolis-Hastings acceptance probabilities. Using the prior over velocities, Equation (4.5) and the property of the delta function, we can compute the posterior as:

$$p(\mathbf{C}|\hat{\mathbf{z}}) = \prod_i \int p(v_{\tau_i}|C_{\tau_i}) e^{-\frac{(\tau_{i+1}-\tau_i)v_{\tau_i}^T v_{\tau_i}}{2\sigma_z^2} + \frac{(z_{\tau_{i+1}}-z_{\tau_i})^T v_{\tau_i}}{\sigma_z^2}} dv_{\tau_i}.$$

Each term in this product is an independent integral over  $v_{\tau_i}$ , which only depends on  $z_{\tau_{i+1}} - z_{\tau_i}, \tau_{i+1} - \tau_i$  and implicitly on the eye state  $C_{\tau_i}$  through the parameters  $d$  and  $\sigma$  in the prior  $p(v_{\tau_i})$ . We can therefore write

$$p(\hat{\mathbf{z}}|\mathbf{C}) = \prod_i I(z_{\tau_{i+1}} - z_{\tau_i}, \tau_{i+1} - \tau_i, d_{C_{\tau_i}}, \sigma_{C_{\tau_i}}),$$

with

$$I(\Delta z, \Delta \tau, d, \sigma) = \frac{1}{2\pi} \iint p(r|d, \sigma) e^{-\frac{\Delta \tau}{2\sigma_z^2} r^2 + \frac{1}{\sigma_z} \Delta z \cdot \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix}} dr d\theta. \quad (4.11)$$

This integral can be evaluated to

$$I(\Delta z, \Delta \tau, d, \sigma) = \frac{2^{\frac{1-d}{2}}}{\sigma^{d+1} \Gamma\left(\frac{d+1}{2}\right)} \left(\frac{\sigma_z^2}{\|\Delta z\|}\right)^{d+1} A\left(d, \frac{\sigma_z^4}{2\|\Delta z\|^2} \left(\frac{1}{\sigma^2} + \frac{\Delta \tau}{\sigma_z^2}\right)\right), \quad (4.12)$$

(For details, see Supplementary. 4.5.1) where

$$A(d, \alpha) = \int_0^\infty s^d e^{-\alpha s^2} I_0(s) ds,$$

with  $I_0$  the modified Bessel function of the first kind. We solve this integral analytically in the limits  $\alpha \rightarrow 0$  and  $\alpha \rightarrow \infty$ , which correspond to  $\|\Delta z\| \rightarrow \infty$  and  $\|\Delta z\| \rightarrow 0$ , respectively. For intermediate  $\alpha$ , we solve the integral numerically.

The details of the MCMC algorithm we use to sample from the posterior  $p(\mathbf{C}|\hat{\mathbf{z}})$  are presented in the Supplementary 4.5.1. The MCMC algorithm returns a set of 40 samples  $\hat{\mathbf{C}}^j$ . On the last iteration, we convert these samples into a probability time series by averaging them. For some applications, we subsequently transform from probabilities to binary values by thresholding at  $\frac{1}{2}$ . This operation minimizes the absolute value cost function

$$\text{Cost}(\hat{\mathbf{C}}, \mathbf{C}) = \sum_{t=1}^T |\hat{C}_t - C_t|.$$

#### 4.2.3.6 Step 4. Estimate the velocity distributions parameters

We infer the global velocity distributions parameters  $\sigma_0, \sigma_1, d_1$  by maximizing  $p(\hat{\mathbf{z}}|\hat{\mathbf{C}}^j, \sigma_0, \sigma_1, d_1, \sigma_z)$  with a grid search for each sample  $\hat{\mathbf{C}}^j$ , and then taking the median across samples. The grid ranges are  $[0.0001, 0.1]$  for  $\sigma_0$  and  $\sigma_1$  and  $[1.1, 5]$  for  $d_1$ .

### 4.2.4 Alternative algorithms

#### 4.2.4.1 Engbert and Kliegl velocity threshold

The Engbert and Kliegl (EK) algorithm starts by averaging the measured eye position time series across a triangular sliding window and differentiating it to obtain a velocity time series. The algorithm detects microsaccades whenever the velocity exceeds a threshold  $\eta_x$  for the horizontal dimension and  $\eta_y$  for the vertical dimension for a sufficiently long duration. The thresholds are adaptively set as a multiple of the standard deviation of the eye movement velocity, using a median-based estimate of the standard deviation.

$$\eta_{x,y} = \lambda \left( \text{median} \left( v_{x,y}^2 \right) - \text{median} \left( v_{x,y} \right)^2 \right).$$

The size of the sliding window, the multiplier  $\lambda$ , and the minimum duration are free parameters set by the user. Of these,  $\lambda$  tends to have the largest effect on the detected microsaccades.

In their original paper, Engbert and Kliegl used a triangular sliding window size of 6 for 500 Hz data, a duration threshold of 12 ms, and a relatively conservative velocity threshold multiplier of  $\lambda = 6$ . This value is used in most subsequent studies. Other studies have

used a more liberal threshold (Engbert and Mergenthaler, 2006). We consider two particular cases with  $\lambda = 3$  and  $\lambda = 6$ , which we will refer to as EK3 and EK6, respectively.

#### *4.2.4.2 Unsupervised clustering*

More recently, Otero-Milan et al. (Otero-Millan et al., 2014a) proposed a threshold-free microsaccade detection method, which we will refer to as OM. It uses an unsupervised clustering algorithm, k-means, to group putative events obtained from the EK algorithm into clusters of microsaccades or drift. The algorithm separates drift and microsaccade events using 3 features: peak velocity, initial acceleration peak and final acceleration peak. Here, we use the implementation provided by Otero-Milan et al., as obtained from their website (Otero-Millan et al., 2014b).

#### **4.2.5 ADHD microsaccade time course analysis**

Microsaccade rates were calculated for each participant for glued task epochs separately for each block. Since the BMD algorithm was not tested for robustness relative to the length of the time series, we applied it separately to the cue periods and separately to the joint delay 1-stimulus-delay 2 periods. For postprocessing, we excluded BMD inferences shorter than 5 ms and eliminated overshoots by excluding what were labeled as new microsaccades but occurred after less than 15 ms relative to a previous microsaccade. The microsaccade rate time courses were calculated over windows of 50 ms, and smoothed by sliding these windows over in 5 ms steps.

#### 4.2.6 Eyelink experimental methods

This study was approved by the New York University Institutional Review Board, in accordance with the Declaration of Helsinki. Five subjects (two female and three male) of median age 26 years old (age range 20 to 36 years) performed the task upon providing informed consent.

**Apparatus.** We displayed stimuli on a 21-inch Sony GDMF520 cathode ray tube (CRT) monitor (resolution:  $1280 \times 960$  pixels, 100 Hz refresh rate). Subjects used a head rest located at approximately 57 cm from the screen. The screen background was gray ( $57 \text{ cd/m}^2$ ). An Apple iMac computer running Matlab 7.1 (MathWorks, Massachusetts, USA) with the Psychtoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and EyeLink extensions (Cornelissen et al., 2002) controlled stimulus presentation and response collection. We obtained monocular recordings of fixational eye movements using a remote infrared video-oculographic system (EyeLink 1000; SR Research) with a 1 kHz sampling rate, precision of 0.01 degrees and average accuracy of 0.25 - 0.5 deg, according to the manual (but see (Holmqvist et al., 2011; Poletti and Rucci, 2016)). We acquired eye position data with the EyeLink software. We set the ‘Heuristic filter’ option OFF to obtain the raw data.

**Procedure.** Subjects performed a delayed-estimation of orientation task, as introduced by (Wilken and Ma, 2004). A trial sequence started with the appearance of a central white fixation cross subtending a visual angle of 0.3 deg, which lasted for 500 ms or until the subject successfully fixated. We defined fixation to be successful when the eye position remains within a 2 deg circle centered at the fixation cross. Next, two stimuli appeared 6 deg to the right and left of the central fixation cross. The stimuli were circularly windowed

gratings with radius 0.35 deg, spatial frequency 2.5 cycles/deg and uniformly drawn orientations. The stimuli stayed on the screen for 11 frames (about 110 ms), followed by a delay period of 1000 ms. If the subject broke fixation at any point during the stimulus or delay period, the trial was aborted and a new trial sequence started. We eliminated these trials from our data set. After the delay period, the subject was probed about one of the locations and they responded by using the mouse to estimate the orientation. More precisely, when the subject moved the mouse, a windowed grating appeared inside that circle. The subject had to rotate it using the mouse to match the orientation of the grating that had been in that location, and then press the space bar to submit a response. The experiment consisted of 8 blocks, each consisting of 60 completed (non-aborted) trials with 30 seconds breaks in between blocks.

#### 4.2.7 DPI Experimental methods

The Dual Purkinje Image eye tracker data were made available by Martina Poletti and Michele Rucci. Their study was approved by the Institutional Review Board of Boston University. The method and data were described in detail in (Cherici et al., 2012) and we summarize them here.

**Apparatus.** Stimuli were presented on a custom-developed system for flexible gaze-contingent display control on a fast-phosphor cathode ray tube (CRT) monitor (Iyama HM204DT) with a vertical refresh rate of 150 Hz. The movements of the right eye were measured with a Generation 6 Dual Purkinje Image (DPI) eye tracker (Fourward Technologies), at 1kHz sampling rate. While most video-based eye trackers only detect the first corneal reflection (Purkinje reflection), DPI eye trackers detect both the first and fourth Purkinje reflections, allowing discrimination between eye rotation and eye trans-

lation movements. The DPI eye tracker has a high precision, of 0.006 deg (Cherici et al., 2012; Crane and Steele, 1985).

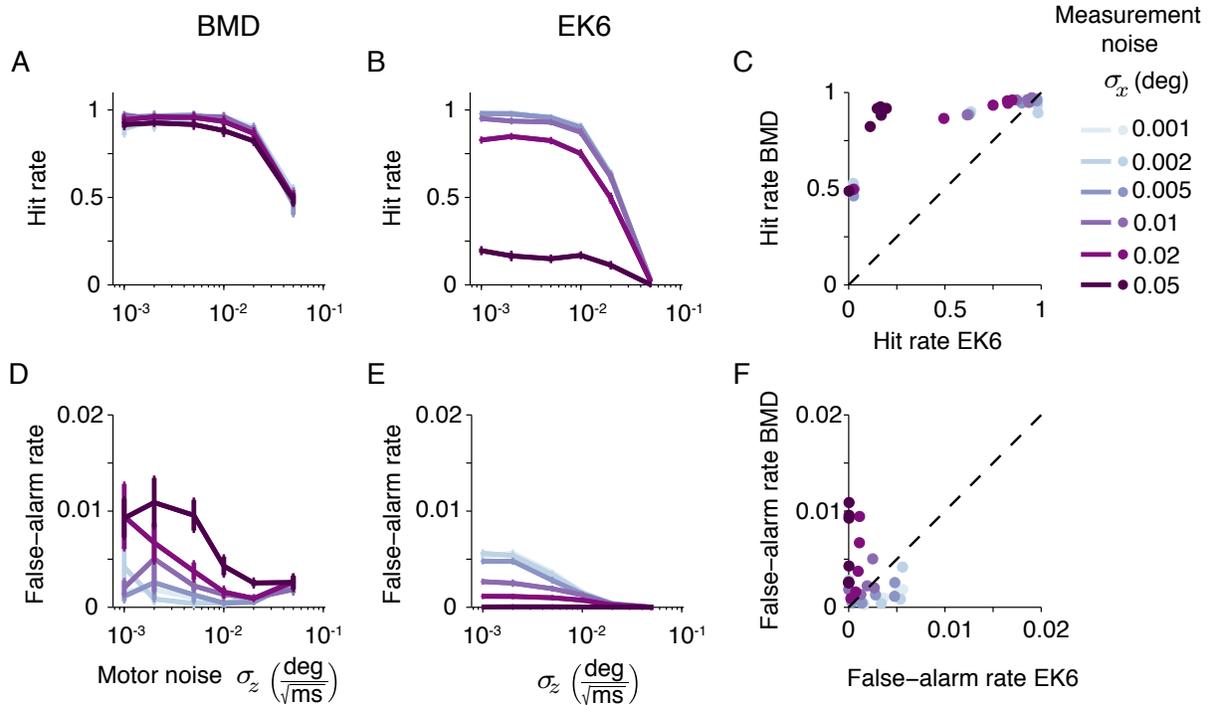
**Procedure.** Subjects observed the screen with the right eye while wearing an eye patch on their left eye. A dental-imprint bite bar and a headrest prevented head movements. Subjects were asked to maintain sustained fixation while looking at a marker displayed on the screen. Two subjects performed the task.

## 4.3 Results

### 4.3.1 Comparison of BMD and other algorithms on simulated data according to our generative model

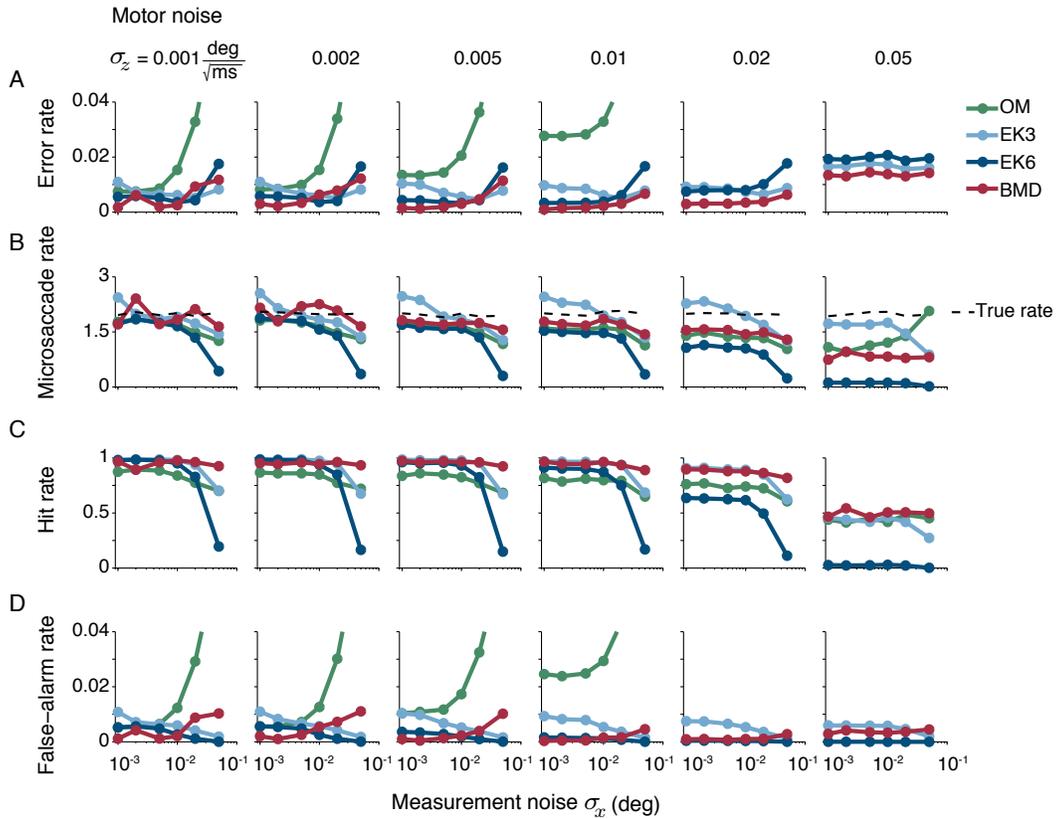
We created 36 data sets with eye position time series of length  $T = 60000$  ms according to the generative model. We created every combination of 6 chosen values of motor noise and 6 values of measurement noise. We fixed the velocity distribution parameters at  $\sigma_0 = 0.3 \frac{\text{deg}}{\text{s}}$ ,  $d_1 = 4.4$  and  $\sigma_1 = 30 \frac{\text{deg}}{\text{s}}$ , to approximate realistic microsaccade kinematics (Engbert, 2006). We inferred the eye state time series with the BMD algorithm and the standard Engbert and Kliegl algorithm, which uses a velocity threshold multiplier of 6, referred to as EK6. After thresholding the BMD inferences, we evaluated their performance in terms of the hit rate (defined as the proportion of 1's correctly identified in the  $\mathbf{C}$  time series) and the false-alarm rate (the proportion of 1's wrongly identified in the  $\mathbf{C}$  time series) (Figure 4.4). While the velocity distribution parameters were not perfectly recovered (Figure 4.19), the BMD hit rates were very high (Figure 4.4A). The hit rate of the BMD algorithm decreases with increased motor noise, as in standard signal detection theory,

but it is remarkably robust to increased measurement noise. By contrast, the hit rate of EK6 is lower and more affected by the noise level. In EK6, the false-alarm rate decreases with increasing noise because the threshold adapts to the noise level. Across the board, BMD has false-alarm rates comparable to EK6, but much higher hit rates, especially at high noise.

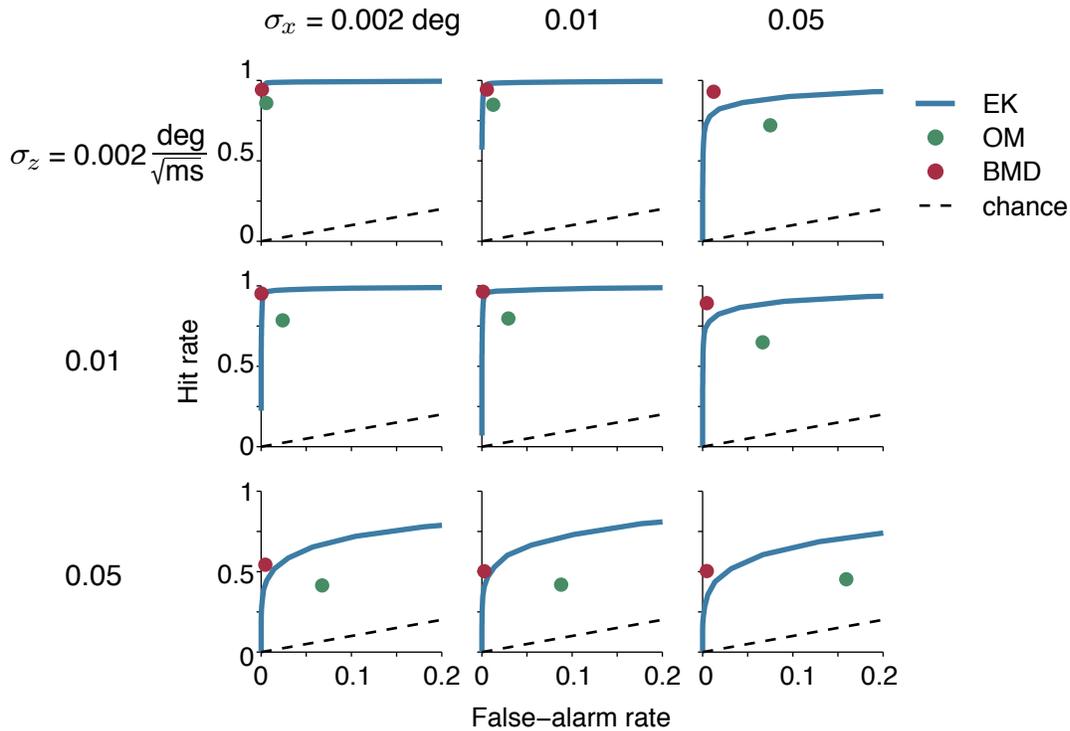


**Figure 4.4** Performance of the BMD and EK6 algorithms on simulated data. **(A)** Hit rates of the BMD algorithm as a function of the motor noise  $\sigma_z$  for several values of measurement noise  $\sigma_x$ . Points and error bars represent means and standard errors across 8 simulated data sets. **(B)** Hit rates of the EK6 algorithm. **(C)** Scatterplot comparing hit rates of both algorithms. Each point corresponds to a different  $\sigma_z$ ,  $\sigma_x$  pair. **(D-F)** Same for false-alarm rates.

For a more comprehensive evaluation, we also compare BMD against OM and an EK variant with a velocity threshold multiplier  $\lambda = 3$  (EK3) (Figure 4.5). As performance metrics, we use the error rate in identifying the eye state at every time point, the number



**Figure 4.5** Performance of several algorithms on simulated data. Colors represent four different algorithms: OM (Otero-Millan et al., 2014a), two versions of EK (Engbert and Kliegl, 2003), and BMD. We evaluate performance with four different metrics: **(A)** error rate, **(B)** microsaccade rate, **(C)** hit rate, and **(D)** false-alarm rate. The motor noise  $\sigma_z$  increases across columns and the measurement noise  $\sigma_x$  increases within each subplot. BMD has the lowest error rates at high noise levels and is the most robust against increases in both  $\sigma_z$  and  $\sigma_x$ . BMD hit rates and microsaccade rates are the most robust against increases in either  $\sigma_z$  or  $\sigma_x$ , without a major increase in false alarm rates.



**Figure 4.6** Performance of the algorithms on simulated data visualized relative to the EK ROC curve. The red and green dots represent the combination of hit rate and false-alarm rate for BMD and OM, respectively. The EK ROC curves were created with different values of the threshold multiplier  $\lambda$ . EK3 and EK6 correspond to points on the curve. For all noise levels tested, including the ones presented here, BMD either outperforms both OM and EK, or matches EK.

of microsaccades per unit time, and the hit and false-alarm rates. BMD has lower error rate than all alternative algorithms in 30 out of 36 noise levels. As in Figure 4.4, the improvement of BMD over alternative algorithms is larger for higher noise. BMD has a hit rate close to 1 in all but the highest level of motor noise, whereas the false-alarm rate is comparable to other algorithms. The BMD algorithm is more robust than all other algorithms: its hit rate and microsaccade rate vary only weakly with increasing measurement noise.

As expected from signal detection theory, there is a trade-off between false alarms and misses in the EK algorithm. EK6 is too conservative, leading to more misses than BMD. However, EK3 is too permissive and has more false alarms. To test whether the EK algorithm with any threshold can match BMD’s performance, we compute a receiver operating characteristic (ROC, Figure 4.6). At low noise, both BMD and EK perform close to perfectly. Overall, BMD outperforms or matches EK at all other noise levels. However, in cases where BMD performance matches EK, the BMD intersects the EK ROC curves for different thresholds at different noise levels. This makes choosing a single best threshold problematic. A more comprehensive evaluation would also vary the threshold we used to collapse the probability inferred by BMD (currently 0.5) and also generate ROC curves for BMD.

#### **4.3.2 Application to real data: BMD inferences on a pilot EyeLink dataset**

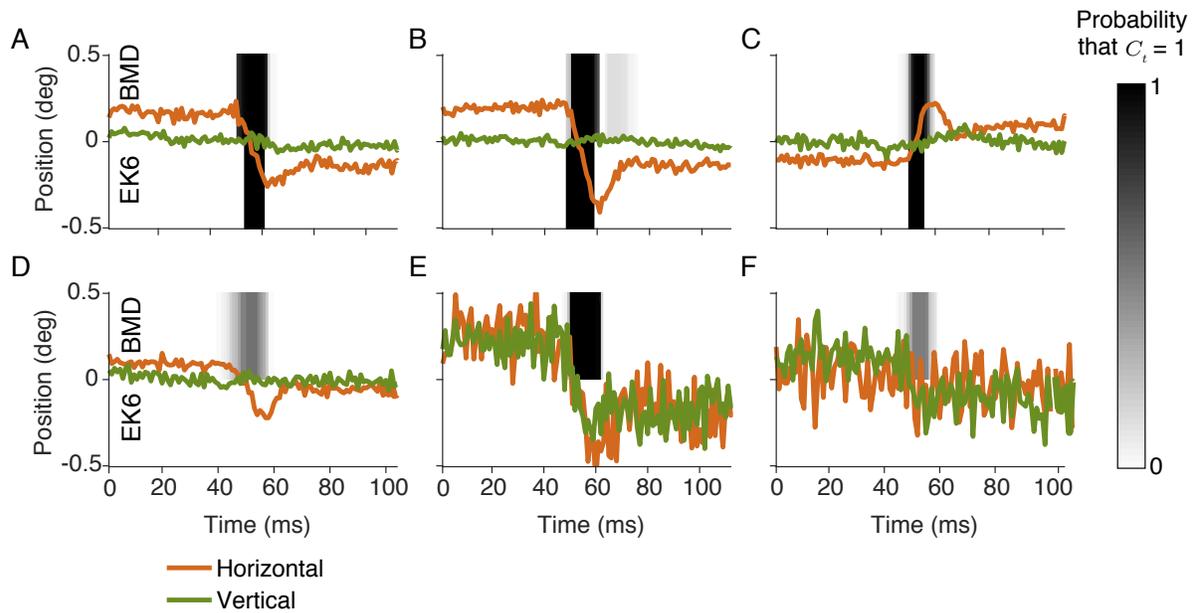
The results on simulated data suggest that BMD recovers microsaccades more faithfully than alternative algorithms, especially at high noise. This confirms that the approximations in our inference algorithm do not significantly impair its performance. However, we created data according to our generative model, so we expected the BMD algorithm to be superior. Next, we apply our algorithm to real eye tracking data measured with two different eye trackers: EyeLink and Dual Purkinje Image (DPI). In Figure 4.7, we show 6 example measured eye position sequences and the inferred change points by BMD and EK6. When the signal-to-noise ratio is high (Figure 4.7 A,B,C), BMD generally infers the same microsaccades as EK6. Additionally, BMD returns a probabilistic judgment of the beginning and end time of the microsaccade. In some cases, BMD detects a small microsaccade immediately after a larger one, in the opposite direction (Figure 4.7 B,C), cor-

responding to the overshoot. For low signal-to-noise data (Figure 4.7 D,E,F), the BMD algorithm tends to detect potential microsaccades that EK6 misses, however, they could be false positives. BMD assigns low confidence to its judgments in ambiguous cases like Figure 4.7 D and F.

The microsaccades detected by BMD have similar kinematics as previously reported (Engbert, 2006) (Section 4.5.3). The inferred velocity and duration distributions of BMD and EK6 are similar, except for the duration cutoff in EK6. Most importantly, the microsaccades detected by BMD follow the main sequence: their amplitude is monotonically related to their peak velocity (Zuber et al., 1965). As in (Engbert and Kliegl, 2003), we consider the approximate recovery of the main sequence relationship evidence for the validity of our detection algorithm. Our algorithm estimates the mean velocity for drift as  $0.1253 \frac{\text{deg}}{\text{s}}$  for all but one subject, and  $22.64 \pm 8.4 \frac{\text{deg}}{\text{s}}$  (mean and standard error across subjects) for microsaccades. These values are in line with literature reports: mean drift velocity of  $0.85 \frac{\text{deg}}{\text{s}}$  in (Poletti et al., 2015) and below  $0.5 \frac{\text{deg}}{\text{s}}$  in (Engbert, 2006; Rolfs, 2009), and mean microsaccade velocity of  $\approx 30 \frac{\text{deg}}{\text{s}}$ .

Overall, BMD detects more microsaccades than EK6 for all 5 subjects (Section 4.5.4). This difference can be dramatic: for two subjects (S3 & S4), EK6 infers no microsaccades at all, whereas BMD infers microsaccade rates up to 2.1 per second. This further suggests that EK6 is too conservative and misses microsaccades when the measurement noise is high. The other algorithms (OM and EK3) are less conservative, but their inferred microsaccade rates vary widely, reinforcing the need for a more principled microsaccade detection algorithm.

Finally, we ask how dependent the microsaccade rate inferred by BMD is on the choice of parameters in the priors over the frequency and duration of microsaccades. We



**Figure 4.7** Inferences of microsaccades by BMD and EK6 on example eye position sequences measured with the EyeLink eye tracker. The black and white shading represents the probability that the eye is in a microsaccade state, with black indicating certainty. Every subplot shows the BMD inference on the top half and the EK6 inference on the bottom half. **(A-C)**. Often, BMD and EK6 infer nearly identical microsaccade sequences. **(D-F)**. BMD infers potential microsaccades that EK6 misses, especially when they are small or noisy.

vary both  $k_0$  and  $k_1$  by an order of magnitude and show that the inferred microsaccade rate is approximately constant (Figure 4.22), making the BMD algorithm robust to the choice of the prior in a plausible range.

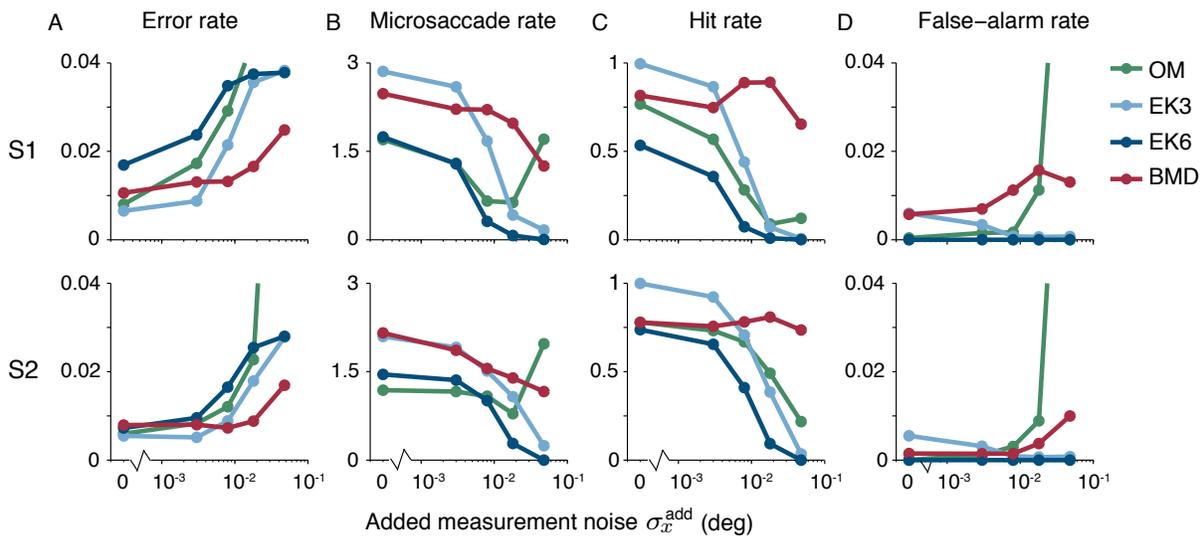
These results suggest that BMD outperforms EK6 in real data. Specifically, BMD detects many plausible microsaccades that EK6 misses, especially when their amplitude is small and the noise is high. However, an alternative interpretation is that BMD detects false positives. We cannot distinguish these possibilities because, in contrast to the simulated data, we do not know the ground truth. In general, we know that all 4 algorithms give different inferences, but without ground truth, we have no way of establishing which one is better.

### 4.3.3 Application to real data: Dual Purkinje Image data

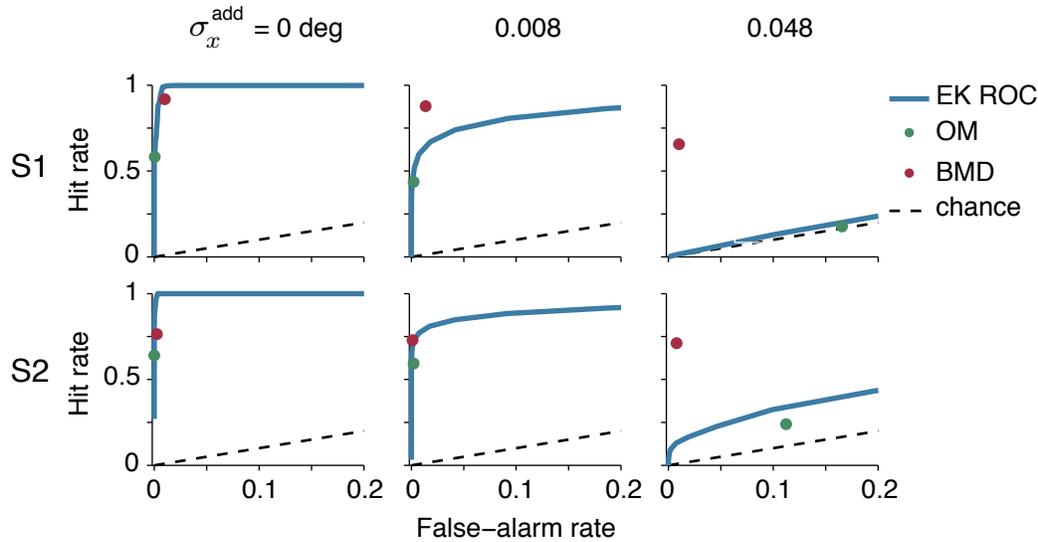
To address this problem, we use another data set, provided by Martina Poletti and Michele Rucci (Cherici et al., 2012). These eye movements were measured with the more precise DPI eye tracker (Cherici et al., 2012; Crane and Steele, 1985). Indeed, BMD infers that the geometric mean of the measurement noise level in DPI data is almost an order of magnitude lower than in EyeLink data (Table 4.2). In simulated data with the same noise level as BMD infers for DPI, all algorithms perform close to perfectly. In view of this high performance, we can treat the microsaccades inferred from the raw DPI data (averaged across algorithms) as ground truth. Our strategy is to artificially add increasing amounts of measurement noise to the raw data, see how much the inference of each algorithm degrades as a result. While we know that the independent measurement noise assumption does not hold for the majority of eye trackers, we still believe this approach would be to some extent informative about algorithms' robustness of inference under noise.

More specifically, this allows us to compare the robustness of the algorithms with an objective metric.

We compare the error rates as well as the microsaccade rates, hit rates, and false-alarm rates between BMD, OM, EK3 and EK6 (Figure 4.8). The BMD algorithm outperforms EK3, EK6, and OM at all except the lowest noise levels. In particular, at measurement noise levels comparable to the ones inferred in EyeLink data (0.02 deg), the error rate for EK6 is 3.22% (averaged across subjects), while BMD achieves 1.48%, a 54% improvement. Note that all algorithms have low error rate, primarily because microsaccades are rare. As in simulated data, we compare BMD to EK with different thresholds by plotting an ROC curve; BMD outperforms EK regardless of its threshold (Figure 4.9). As mentioned in the case of simulated data, a more comprehensive evaluation would also generate ROC curves for BMD and better understand its hits and false-alarm rate trade-offs.



**Figure 4.8 Performance of the algorithms on DPI data.** We took DPI data from two subjects (rows), collected by (Cherici et al., 2012) and artificially added measurement noise to the eye position traces. Colors represent algorithms. BMD shows the highest robustness to adding measurement noise; specifically, error rates are lowest and hit rates tend to stay the same.



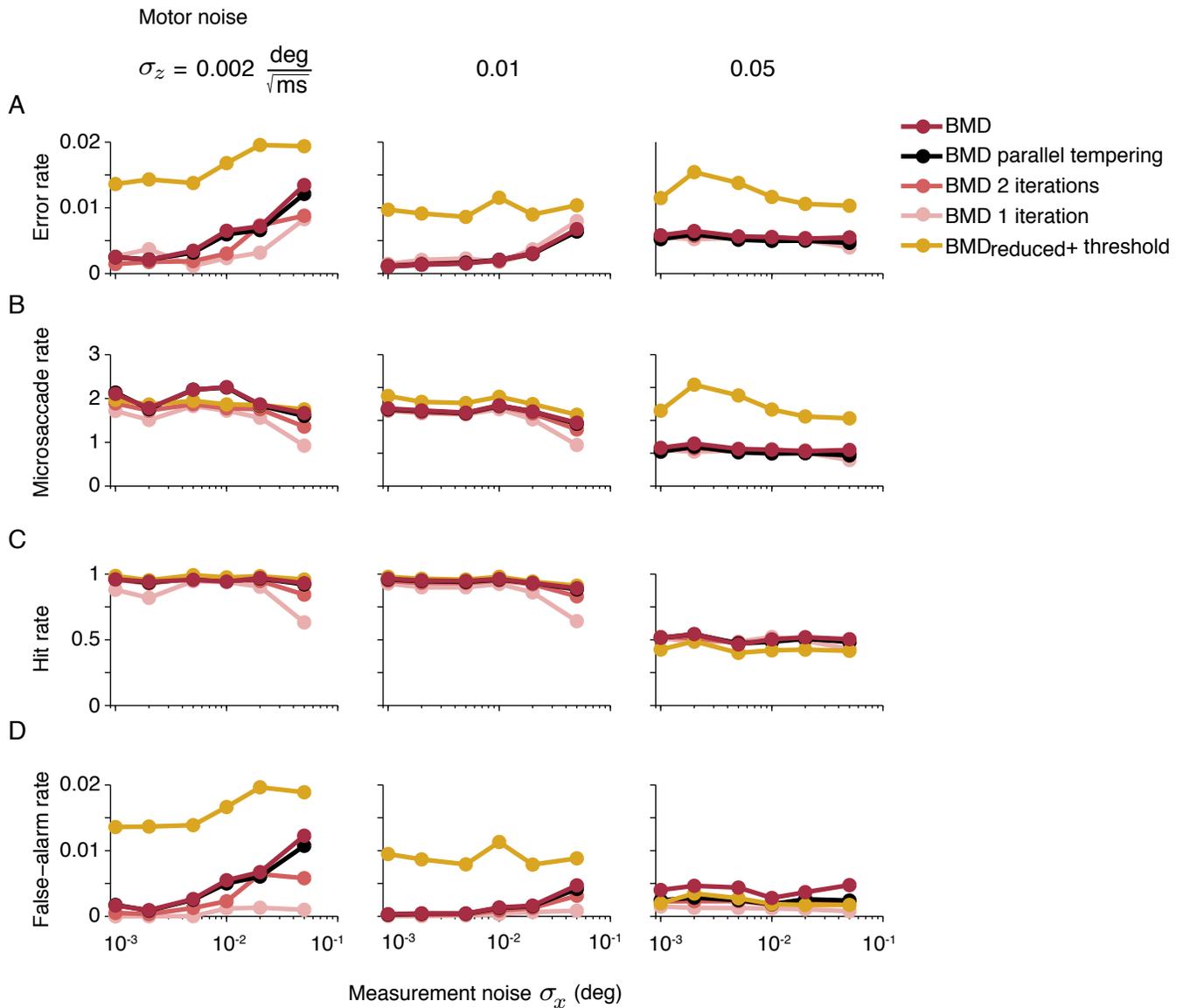
**Figure 4.9** Performance of the algorithms on DPI data with added noise relative to the EK ROC curve. We show the two subjects, S1 and S2, as in Figure 4.8. The level of the added measurement noise varies across columns. The EK ROC curves were created with different values of the threshold multiplier  $\lambda$ . EK3 and EK6 correspond to points on the curve. As more measurement noise is added, BMD outperforms EK and OM by larger amounts.

#### 4.3.4 Variants of BMD

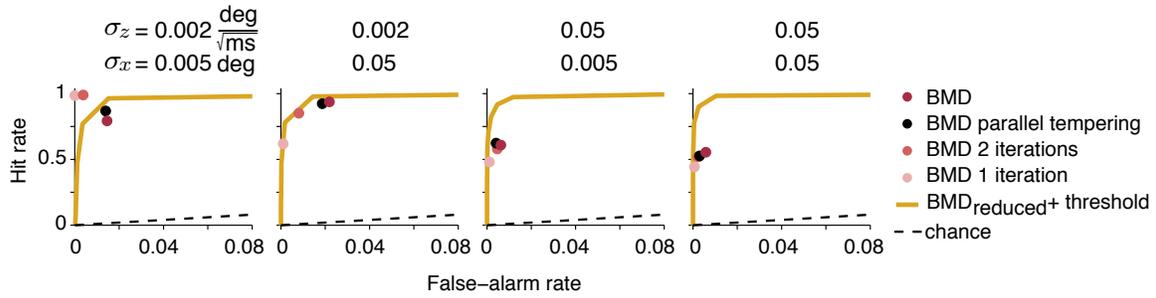
A common risk in Monte Carlo methods is that the samples aggregate near potential local maxima of the posterior, and miss the global maximum. One method to mitigate this problem is parallel tempering (Earl and Deem, 2005; Newman and Barkema, 1999), albeit at increased computational cost. BMD with parallel tempering does not significantly outperform BMD neither in simulated data (Figure 4.10, Figure 4.11) nor in real DPI data with added noise (Figure 4.12), suggesting that the posterior probability landscape did not contain many local maxima. To investigate which components of our method are necessary for its performance, we compare BMD against three reduced variants. We obtain the first version by reducing the number of iterations in the approximate inference method from 6 to 2. The second version has only 1 iteration, which is equivalent to applying a Kalman

smoother to obtain  $\hat{\mathbf{z}}$  from  $\mathbf{x}$ , then sampling from  $p(\mathbf{C}|\hat{\mathbf{z}})$ .

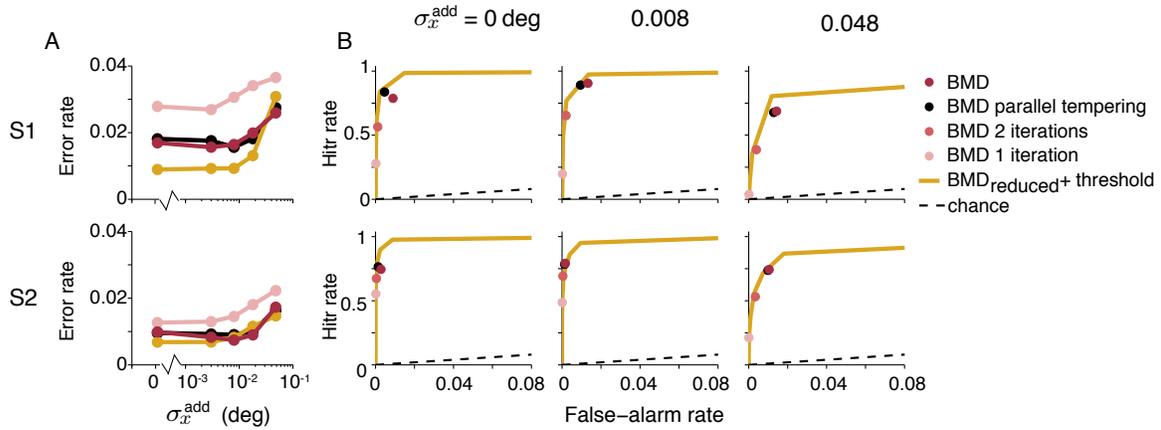
Finally, a third version,  $\text{BMD}_{\text{reduced}} + \text{threshold}$ , starts with Steps 0 to 2 of the BMD algorithm. However, instead of sampling from the posterior  $p(\mathbf{C}|\hat{\mathbf{z}})$  in Step 3, it estimates  $\mathbf{C}$  by applying a Kalman smoother (after the Kalman smoother of Step 2) to  $\hat{\mathbf{z}}$  to obtain a smoothed eye position time series, differentiating that to obtain eye velocities, and thresholding the velocity time series (Figure 4.13). We fix the window size of the Kalman smoother to 5.32 ms and use a threshold which scales linearly with the inferred motor noise level:  $\text{threshold} = a\hat{\sigma}_z + b$ , with  $a = 32\sqrt{\text{s}}^{-1}$  and  $b = 1\frac{\text{deg}}{\text{s}}$ . We chose these values to approximately match the output of BMD and  $\text{BMD}_{\text{reduced}} + \text{threshold}$  in real and simulated data. This method performs about as well as the full inference algorithm. However, it is unprincipled, does not return a probabilistic estimate, and cannot be directly extended to more sophisticated generative models. Additionally, while it is plausible to wonder whether the Kalman smoother is the main ingredient responsible for the increased performance of BMD, we caution that our the parameter ranges tested here are far from exhaustive and thus we cannot assume without further testing that the performance of  $\text{BMD}_{\text{reduced}} + \text{threshold}$  would generalize to datasets with other parameters.



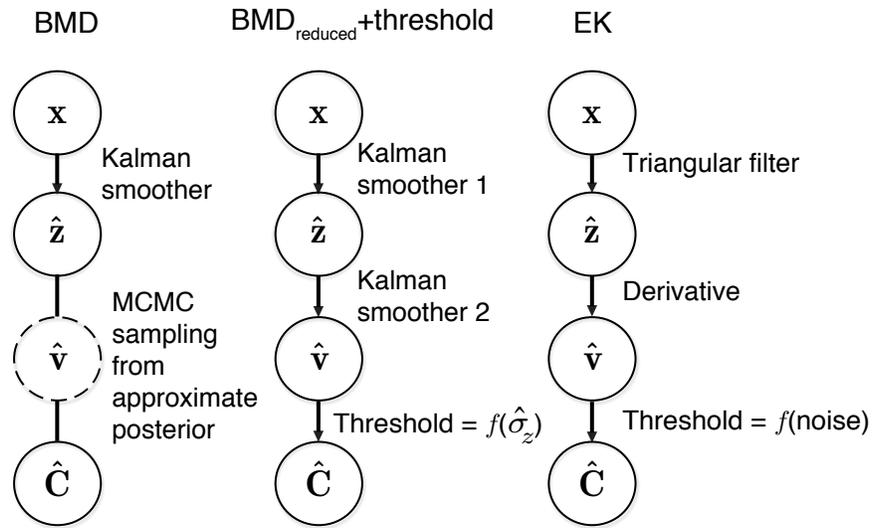
**Figure 4.10 Performance of BMD variants on simulated data.** The variants we examine are BMD with parallel tempering, BMD with fewer iterations (2 and 1) and a reduced variant of BMD with a threshold (BMD<sub>reduced</sub> + threshold). In the latter model, the threshold is dependent on motor noise through the equation:  $\text{threshold} = 32\sqrt{5}^{-1}\hat{\sigma}_z + 1\frac{\text{deg}}{\text{s}}$ , chosen because it gave the lowest error rates on DPI data. The motor noise  $\sigma_z$  increases across columns and the measurement noise  $\sigma_x$  increases within each subplot. BMD with parallel tempering is only a slight improvement over BMD, while BMD performs slightly better than BMD with 2 and 1 iterations. BMD<sub>reduced</sub> + threshold only performs comparably with BMD under high motor and measurement noise.



**Figure 4.11** Performance of BMD variants on simulated data visualized relative to the BMD<sub>reduced</sub> + threshold ROC curves. Hit rates with false-alarm rates points for BMD variants and ROC curves for BMD<sub>reduced</sub> + threshold for several values of the threshold multiplier. In contrast to Figure 4.10, where we choose one threshold, here we see that the BMD variants points are on the ROC curves at low noise (first two subplots). However, as the motor and measurement noise increase (last two subplots), the full ROC curve can reach higher performance than the BMD variants algorithms.



**Figure 4.12** Performance of BMD variants on DPI data to which we add measurement noise. We show the two subjects, S1 and S2, as in Figure 4.8. We measure performance on the same metrics as before. For brevity, we show in (A) the error rates with fixed threshold. In (B), we show the hit rates with false alarm rates for the BMD variants algorithms relative to the BMD<sub>reduced</sub> + threshold ROC curve. Adding parallel tempering to BMD makes little difference. Using fewer iterations negatively affects the hit rate. BMD<sub>reduced</sub> + threshold gives slightly lower error rates and seems to match the performance of BMD.

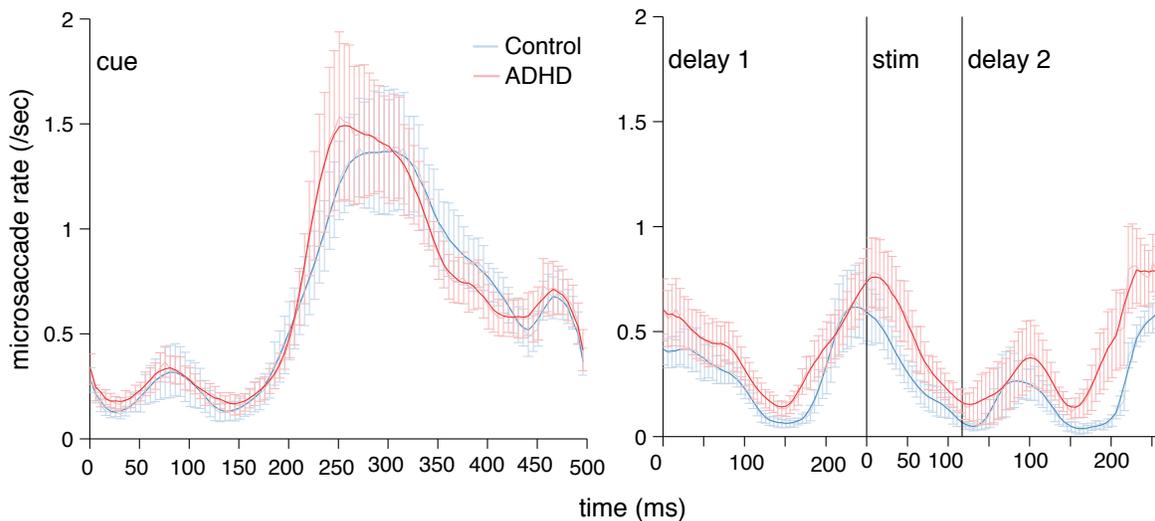


**Figure 4.13 Schematic comparison of microsaccade detection algorithms.** All algorithms first perform a filtering operation to eliminate the noise from the measured eye position time series  $\mathbf{x}$ . The EK algorithm removes noise with a heuristically chosen filter; in contrast, BMD and BMD<sub>reduced</sub> + thresh use a Kalman smoother, which optimally eliminates measurement noise in our generative model. The EK algorithm estimates the eye state time series by taking the derivate of the eye position to yield the eye velocity time series, and thresholding those velocities. Instead, the BMD algorithm marginalizes over velocity and samples from the posterior distribution over eye states. BMD<sub>reduced</sub> + threshold uses a second Kalman smoother to eliminate some of the motor noise and ultimately uses a velocity threshold which depends on the motor noise.

### 4.3.5 Application to real data: ADHD and Controls EyeLink dataset

We present the microsaccade time courses of the task presented in Section 3.2.1, averaged across trials, conditions and participants within a group. We look at the cue period and separately at the joint delay 1, stimulus and delay 2 periods. We recognize features of the expected microsaccade rate signature. The ADHD and Control traces of mean  $\pm$  sem overlap to a good extent (Figure 4.14). It is likely that a potential difference between ADHD and Controls across a time interval that might turn out statistically significant will not survive a multiple-comparisons correction.

#### 4.3.5.1 Microsaccade time courses



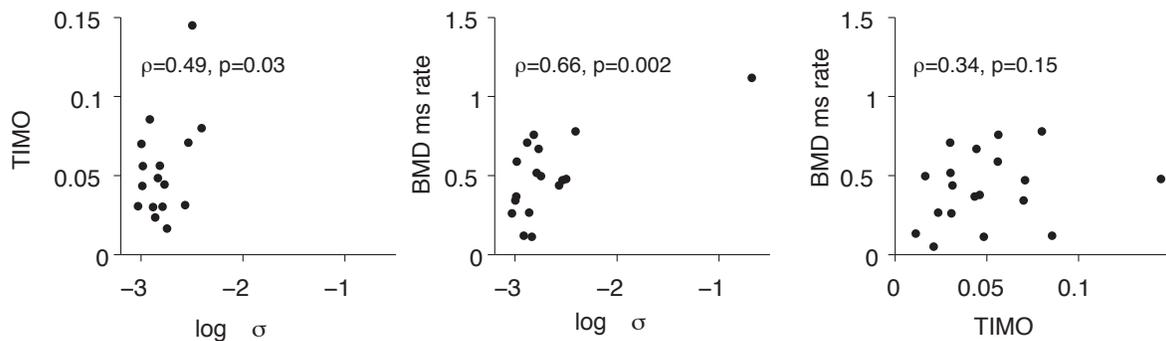
**Figure 4.14** Microsaccade time course. Lines and error bars represent means and standard errors of means across participants within a group.

While in the cue period the rebound visually dominates, in the joint delay 1 - stimulus - delay 2 period the fast presentation of either fixation dot or stimuli yields suppress-

sion. Suppression around stimulus onset can facilitate encoding of information. Could it be that ADHD participants have reduced ability to suppress oculomotor behavior during this period? The microsaccade rates during this joint interval have the mean  $\pm$  sem  $0.35 \pm 0.06$  for Controls and  $0.55 \pm 0.09$  for ADHD. A Wilcoxon rank-sum test did not find a significant difference between these two groups ( $p = 0.12$ ). We note that there seem to be large within group variability, and examine next individual differences.

We collapse across both groups and attempt to correlate the individuals' microsaccade rates with our main task metrics from Chapter 3, perceptual variability and TIMO. Figure 4.15 presents these correlations. Note that the correlation between perceptual variability and TIMO is slightly different than the one in Chapter 3, as here we only have half of the participants. Indeed, microsaccade rate is correlated with the perceptual variability parameter, but not with TIMO, as expected if it were to be related to the precision of stimulus encoding, but not with response selection.

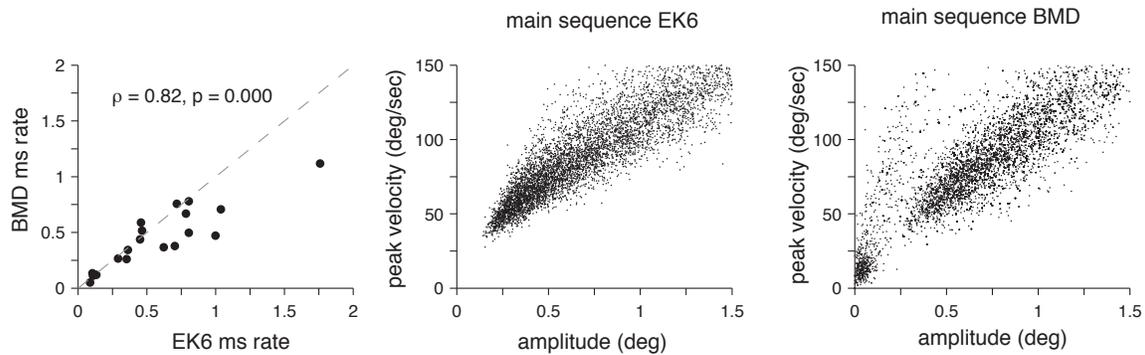
#### 4.3.5.2 *Microsaccade rate correlations with task parameters*



**Figure 4.15** Spearman correlations of microsaccade rate, perceptual variability  $\sigma$  and TIMO. Microsaccade rate correlates with  $\sigma$  but not with TIMO.

#### 4.3.5.3 BMD vs EK6 main sequences and comparison

As a sanity check, we verify these correlations also with the microsaccade rate as inferred by EK6 and get very similar correlations: with perceptual variability  $\rho = 0.69, p = 0.001$  and with TIMO,  $\rho = 0.34, p = 0.15$ . Thus our results are not critically dependent on inferring microsaccades with BMD. Indeed, BMD and EK inferred microsaccade rates are highly correlated (Figure 4.16). This figure also shows the main sequence for the microsaccades inferred by both BMD and EK6, though with some notable differences.



**Figure 4.16** Microsaccades inferred by BMD and EK6 are highly correlated and display the main sequence monotonic relationship.

## 4.4 Discussion

We developed a Bayesian algorithm for detecting microsaccades among drift/tremor; it returns probabilistic rather than binary judgments. Given our assumptions about the statistical process generating a measured eye position time series, this algorithm is optimal. BMD has lower error rates than the algorithms proposed by (Engbert and Kliegl, 2003) and (Otero-Millan et al., 2014a), especially at high noise. This is a particularly useful feature given the relatively high measurement noise of current infrared eye trackers. However, a hybrid between BMD and velocity threshold algorithms,  $\text{BMD}_{\text{reduced}} + \text{threshold}$ , can sometimes approach BMD's performance.

In our model, microsaccades are defined through prior probability distributions over velocity and duration that are different from those for drift/tremor (Figure 4.2). This definition contrasts with the more common one that uses an arbitrary velocity threshold. The BMD algorithm (and the actual code) allows researchers to easily build in their own prior beliefs and state clearly which of their findings depend on those beliefs.

We designed the BMD algorithm for offline analysis of eye tracker data. An online detection method, for example for closed-loop experiments that require real-time detection of microsaccades, such as in (Chen and Hafed, 2013; Yuval-Greenberg et al., 2014), would require a modified inference algorithm. If it is crucial to detect microsaccades online, we recommend using the  $\text{BMD}_{\text{reduced}} + \text{threshold}$ , with a Kalman filter (only the forward filter) instead of the Kalman smoother.

We designed and tested BMD for detecting microsaccades in fixational eye movement data obtained under head-fixed conditions, where the fixation point does not move. Would

the algorithm readily apply to other kinds of eye movement data? First, head-free recordings are sometimes used in order to better mimic naturalistic conditions (Benedetto et al., 2011; Martinez-Conde et al., 2006; Poletti et al., 2015). In theory, our algorithm is suitable for inferring microsaccades in head-free recordings. However, studies have reported higher velocities for drift in head-free fixation (Poletti et al., 2015; Skavenski et al., 1979), for example, on average 1.5 deg/sec for head-free vs 0.85 deg/sec for head-fixed in (Poletti et al., 2015). Therefore, we expect the velocity distributions presented in Figure 3B to be less separable, which in turn would impair microsaccade detection. Second, our algorithm is not immediately applicable to smooth pursuit, in which the eye continuously tracks the motion of an object. (Santini et al., 2016) used a Bayesian classification algorithm to separate drift, saccades and smooth pursuit based on features derived from the eye position data, but this algorithm does not have a generative model of the entire time series. In our approach, we could amend our generative model to include a third, “smooth pursuit”, state with different duration and velocity distributions, but this would require a more complex inference algorithm.

#### 4.4.1 Caveats of the BMD algorithm

BMD outperforms the alternative algorithms in simulated and real data, on average. However, BMD sometimes makes idiosyncratic mistakes. In simulated data with low noise, for some visually salient microsaccades, BMD incorrectly identifies a microsaccade as mostly drift, with very short microsaccades immediately before and after (see Section 4.5.5). This mistake happened 17 times in  $36 \times 8$  simulations. This particular mistake coincides with instances when the algorithm overestimates  $\sigma_0$  and underestimates  $d_1$ , which makes the drift and microsaccade velocity distributions less separable. We could

solve the incorrect microsaccade inference with some post-hoc processing; however, this would introduce arbitrariness. Instead, we accept this as a failure mode of our algorithm: rare, exclusively at low noise, and easily detectable.

While we have not characterized in detail the performance of BMD on EyeLink filtered datasets, we believe it might suffer from the idiosyncracies found in the low-noise regime of the simulated data. BMD's broader applicability may be currently limited due to the fact that several researchers have datasets of eye movement traces that have been filtered online with the "Heuristic filter" (Cornelissen et al., 2002; Stampe, 1993b). There are reasons for more widespread recordings with the "Heuristic filter" option OFF: better user control and the ability to apply offline an optimal filter such as the Kalman smoother.

#### 4.4.2 Conceptual extensions of the BMD algorithm

The inferred microsaccades depend on assumptions in our generative model, which are simplistic and incorrect. We can flexibly adjust these assumptions in the generative model and modify the BMD algorithm accordingly.

**Correlated state durations.** Our generative model assumes that the durations over which the eye remains in either state are independent. We can relax this assumption by changing the duration prior; this does not affect the likelihood.

**Binocular data.** Our algorithm is designed to operate on monocularly recorded eye position time series, but can be applied independently to each eye trace from a binocular dataset. Would be interested to see how BMD compares to the alternative algorithms when considering binocular recordings and imposing the condition of simultaneity of occurrence of microsaccades in both eyes, as it is sometimes practiced (Engbert, 2006; Engbert

and Mergenthaler, 2006; Otero-Millan et al., 2008). Additionally, BMD can be in theory extended to binocular data simply by changing all position and velocity vectors from 2D to 4D and adjusting the noise covariance matrices.

**Tremor and saccades.** We can add tremor (low-amplitude high-frequency oscillations, (Ratliff and Riggs, 1950)) or saccades as additional states in our generative model, given statistical descriptions of these processes. However, it has been argued that microsaccades and saccades are produced by the same process (Hafed and Krauzlis, 2012; Otero-Millan et al., 2008; Zuber et al., 1965).

**Microsaccade dynamics.** Our generative model assumes that the eye velocity is constant throughout each microsaccade or drift state, resulting in linear microsaccade trajectories. However, real microsaccades, such as the ones in Figure 4.1 and Figure 4.7, have a smooth velocity profile, for which specific shapes have been proposed (Abadi and Gowen, 2004). We could incorporate a template for the characteristic temporal profile of microsaccades into our generative model, which would require only minor changes to the inference algorithm.

**Correlated measurement noise.** We assumed that the measurement noise is uncorrelated across time, which allowed us to estimate the eye position using a Kalman smoother (Step 2). We can incorporate noise correlations into our generative model, if we replace the Kalman smoother in the inference algorithm with a Gaussian process estimator.

#### 4.4.3 Microsaccades and perceptual variability in ADHD

While preliminary, our results suggest a potential oculomotor correlate of perceptual variability in less effective microsaccadic suppression around stimulus onset. Such

microsaccadic suppression has been linked to a few measures of performance (Betta and Turatto, 2006; Bonnef et al., 2015; Hafd et al., 2011). Related to ADHD, our results are in line with the results of (Dankner et al., 2017), in which ADHD participants presented a less effective marker of temporal expectation, specifically less effective suppression in microsaccade rate around stimulus onset in a fixed ISI condition relative to a variable ISI condition. In future analyses, we will check the extent and time courses in which the orientations of the microsaccades in the cue period follow the direction in which spatial attention was cued. In addition, we will perform more thorough investigations of microsaccade amplitude, velocity and orientation distributions across each task period are needed.

Relatedly, a previous study in controls found an effect of fatigue on peak microsaccade velocity and drift mean velocity (DiStasi et al., 2013). It would be interesting to see if microsaccade velocities present different patterns of change across time in ADHD versus Controls. Drift velocity might be hard to interpret as it could be confounded with head motion, which was only to a limited extent constrained with the head rest. In addition to influencing drift velocity, we suspect that head motion might also influence our estimate of the motor noise parameter  $\sigma_z$ , assumed to be from the eye. Examining these metrics likely influenced by head motion could be of particular interest in our ADHD dataset since there is some evidence that head motion might be related to impulsivity symptoms (Kong et al., 2014).

## 4.5 Supplementary

### 4.5.1 Mathematical details of the BMD algorithm

*Computation of the likelihood*

We plug in the distributions from Equations (4.3), (4.5) and (4.6) into the likelihood Equation (4.7):

$$\begin{aligned}
p(\mathbf{x}|\mathbf{C}) &= \iint p(\mathbf{x}|\mathbf{z})p(\mathbf{z}|\mathbf{C}, \mathbf{v})p(\mathbf{v}|\mathbf{C})d\mathbf{v}d\mathbf{z} \\
&= \iint \prod_t \mathcal{N}(x_t; z_t, \Sigma_x)\mathcal{N}(z_t; z_{t-1} + v_t, \Sigma_z) \prod_i \left[ p(v_{\tau_i}|C_{\tau_i}) \prod_{t=\tau_i+1}^{\tau_{i+1}-1} \delta(v_t - v_{t-1}) \right] d\mathbf{v}d\mathbf{z} \\
&\propto \iint \prod_t e^{-\frac{(x_t-z_t)^\top(x_t-z_t)}{2\sigma_x^2}} e^{-\frac{(\Delta z_t-v_t)^\top(\Delta z_t-v_t)}{2\sigma_z^2}} \prod_i \left[ p(v_{\tau_i}|C_{\tau_i}) \prod_{t=\tau_i+1}^{\tau_{i+1}-1} \delta(v_t - v_{t-1}) \right] d\mathbf{v}d\mathbf{z},
\end{aligned}$$

where  $\Delta z_t = z_t - z_{t-1}$ . We then expand the product  $\frac{(\Delta z_t-v_t)^\top(\Delta z_t-v_t)}{2\sigma_z^2}$  and gather terms that depend on  $v_t$  into the second integral:

$$p(\mathbf{x}|\mathbf{C}) \propto \iint \prod_t e^{-\frac{(x_t-z_t)^\top(x_t-z_t)}{2\sigma_x^2} - \frac{\Delta z_t^\top \Delta z_t}{2\sigma_z^2}} \prod_i \left[ p(v_{\tau_i}|C_{\tau_i}) \prod_{t=\tau_i+1}^{\tau_{i+1}-1} e^{-\frac{-v_t^\top v_t + 2\Delta z_t v_t}{2\sigma_z^2}} \delta(v_t - v_{t-1}) \right] d\mathbf{v}d\mathbf{z}.$$

The delta function will collapse the integral over the time series  $\mathbf{v}$  to the integral over the change points  $v_{\tau_i}$ , yielding Equation (4.8).

### *Kalman smoother*

The goal of Step 2 in the approximate inference algorithm is to maximize the integrand of the first term of Equation (4.8):

$$e^{-\frac{(z_t - z_{t-1})^T (z_t - z_{t-1})}{2\sigma_z^2} - \frac{(x_t - z_t)^T (x_t - z_t)}{2\sigma_x^2}}.$$

This integrand can be interpreted as the likelihood of a stochastic process with update equations

$$\begin{aligned} z_t &= z_{t-1} + \zeta_t \\ x_t &= z_t + \xi_t, \end{aligned}$$

where  $\zeta$  and  $\xi$  represent independent Gaussian noise with standard deviation  $\sigma_z$  and  $\sigma_x$ , respectively. These equations represent a special case of the Kalman update equations (Kalman, 1960; Welch and Bishop, 2006); therefore, the maximum-likelihood estimate of  $\mathbf{z}$  given  $\mathbf{x}$  is a special case of a Kalman smoother. In a Kalman filter, the goal would be to predict a future state  $z$  based on the observations  $x$  so far. However, since we have access to the entire time series, the correct inference of  $z$  is given by a Kalman smoother. This can be implemented using the RTS algorithm (Rauch et al., 1965; Terejanu, 2008). In our case, this takes the form of a Kalman filter forward in time, followed

by another Kalman filter backward in time. The forward Kalman filter is:

$$\begin{aligned}
 K_t &= \frac{P_{t-1} + \sigma_z^2}{P_{t-1} + \sigma_z^2 + \sigma_x^2} \\
 \hat{y}_t &= \hat{y}_{t-1} + K_t(x_t - \hat{y}_{t-1}) \\
 P_t &= (1 - K_t)(P_{t-1} + \sigma_z^2).
 \end{aligned} \tag{4.13}$$

In these equations,  $P_t$  is the variance of the posterior over  $\hat{y}_t$ , and  $K_t$  is the Kalman gain. If  $P_t$  is large, we expect large changes in the states so we need to be able to update the estimates with new incoming measurements  $x_t$ . Higher weighting of the incoming measurements is achieved with increased Kalman gains. However, if the measurement noise  $\sigma_x$  is high, the observation  $x_t$  is less reliable and the Kalman gain will decrease accordingly, weighing the observation less when estimating the state  $\hat{y}_t$ .

The variance of the posterior  $P_t$  does not depend on the observations  $\mathbf{x}$ , but only on  $\sigma_z$  and  $\sigma_x$ , and on  $P_{t-1}$  through a recurrence relation that follows from Equation (4.13):

$$P_t = \frac{(P_{t-1} + \sigma_z^2)\sigma_x^2}{P_{t-1} + \sigma_z^2 + \sigma_x^2}.$$

This recurrence relation defines  $P_t$  at each time point given a choice for  $P_0$ , the variance of the prior over the first time point. The choice of  $P_0$  affects the variance  $P_t$  at early times, but not for  $t \gg \frac{\sigma_x^2}{\sigma_z^2}$ , because the recurrence relation converges. At convergence,  $\lim_{t \rightarrow \infty} P_{t-1} = \lim_{t \rightarrow \infty} P_t = P$ . Plugging this into the forward update equations yields the quadratic equation  $P^2 + P\sigma_z^2 - \sigma_z^2\sigma_x^2 = 0$  with the valid solution:

$$P = \frac{-\sigma_z^2 + \sqrt{\sigma_z^4 + 4\sigma_z^2\sigma_x^2}}{2}.$$

We choose  $P_0 = P$ , which implies  $P_t = P$  for each time point. Therefore, the Kalman gain  $K_t$  is also constant across time. Plugging this Kalman gain into Equation (4.9) allows us to express the state estimation equation for  $\hat{y}_t$  in terms of the previous estimation  $\hat{y}_{t-1}$ , current observation,  $x_t$  and the process and noise standard deviations:

$$\hat{y}_t = \hat{y}_{t-1} + \frac{\sigma_z^2 + \sqrt{\sigma_z^4 + 4\sigma_z^2\sigma_x^2}}{2\sigma_x^2 + \sigma_z^2 + \sqrt{\sigma_z^4 + 4\sigma_z^2\sigma_x^2}}(x_t - \hat{y}_{t-1}).$$

We denote  $R = \frac{\sigma_x}{\sigma_z}$  and then get:

$$\hat{y}_t = \hat{y}_{t-1} + \frac{1 + \sqrt{1 + 4R^2}}{1 + 2R^2 + \sqrt{1 + 4R^2}}(x_t - \hat{y}_{t-1}).$$

Next, we apply a second Kalman filter, backwards in time, to the output of the first filter  $\hat{y}$  to yield the estimated eye position  $\hat{z}$  (Rauch et al., 1965; Terejanu, 2008). We initialize the eye position at the end of the time series  $\hat{z}_T$  to be equal to  $\hat{y}_T$ , and again set the prior variance of  $\hat{z}_T$  equal to the asymptotic limit  $P$ . This backwards filter has a different Kalman gain than the first filter; the RTS update equations in our case yield  $K = \frac{P}{P + \sigma_z^2}$ . We can rewrite it, and thus the update equation for eye position Equation (4.10) becomes:

$$\hat{z}_t = \hat{y}_t + \frac{\sqrt{1 + 4R^2} - 1}{\sqrt{1 + 4R^2} + 1}(\hat{z}_{t+1} - \hat{y}_t).$$

*From Equation (4.11) to Equation (4.12)*

Plugging Equation (4.4) into Equation (4.11) , we get:

$$\begin{aligned}
I(\Delta z, \Delta\tau, d, \sigma) &= \frac{2^{\frac{1-d}{2}}}{2\pi\Gamma(\frac{d+1}{2})\sigma^{d+1}} \iint r^d e^{-\frac{r^2}{2\sigma^2}} e^{-\frac{\Delta\tau}{2\sigma_z^2}r^2 + \frac{1}{\sigma_z^2}\Delta z \cdot (r \cos\theta)} dr d\theta \\
&= \frac{2^{\frac{1-d}{2}}}{\Gamma(\frac{d+1}{2})\sigma^{d+1}} \int_0^\infty r^d e^{-r^2\left(\frac{1}{2\sigma^2} + \frac{\Delta\tau}{2\sigma_z^2}\right)} \left[ \frac{1}{2\pi} \int_0^{2\pi} e^{\frac{r}{\sigma_z^2}\|\Delta z\| \cos(\theta - \frac{\pi}{2})} d\theta \right] dr \\
&= \frac{2^{\frac{1-d}{2}}}{\Gamma(\frac{d+1}{2})\sigma^{d+1}} \int_0^\infty r^d e^{-r^2\left(\frac{1}{2\sigma^2} + \frac{\Delta\tau}{2\sigma_z^2}\right)} I_0\left(\frac{r}{\sigma_z^2}\|\Delta z\|\right) dr, \tag{4.14}
\end{aligned}$$

where  $I_0$  is the modified Bessel function of the first kind of order zero. Next, we change variables from  $r$  to  $s = \frac{r}{\sigma_z^2}\|\Delta z\|$  and write the final form of the integral as:

$$I(\Delta z, \Delta\tau, d, \sigma) = \frac{2^{\frac{1-d}{2}}}{\Gamma(\frac{d+1}{2})\sigma^{d+1}} \left( \frac{\sigma_z^2}{\|\Delta z\|} \right)^{d+1} \int_0^\infty s^d e^{-s^2\left(\frac{\sigma_z^4}{2\|\Delta z\|^2}\left(\frac{1}{\sigma^2} + \frac{\Delta\tau}{\sigma_z^2}\right)\right)} I_0(s) ds.$$

This expression shows that we can calculate  $I(\Delta z, \Delta\tau, d, \sigma)$  by evaluating the integral

$$A(d, \alpha) = \int_0^\infty s^d e^{-\alpha s^2} I_0(s) ds,$$

and plugging in  $\alpha = \frac{\sigma_z^4}{2\|\Delta z\|^2} \left( \frac{1}{\sigma^2} + \frac{\Delta\tau}{\sigma_z^2} \right)$ .

Unfortunately, this integral appears to have no general analytic solution. However, in the limit of small or large  $\alpha$ , we can replace the Bessel function by asymptotic approximations and solve the resulting integrals. Specifically, we define upper and lower bounds  $\alpha_\infty(d)$  and  $\alpha_0(d)$ . For  $\alpha < \alpha_0(d)$ , we use the large- $s$  approximation to the Bessel function (Abramowitz and Stegun, 1965),  $I_0(s) \approx \frac{e^s}{\sqrt{2\pi s}}$ , so that

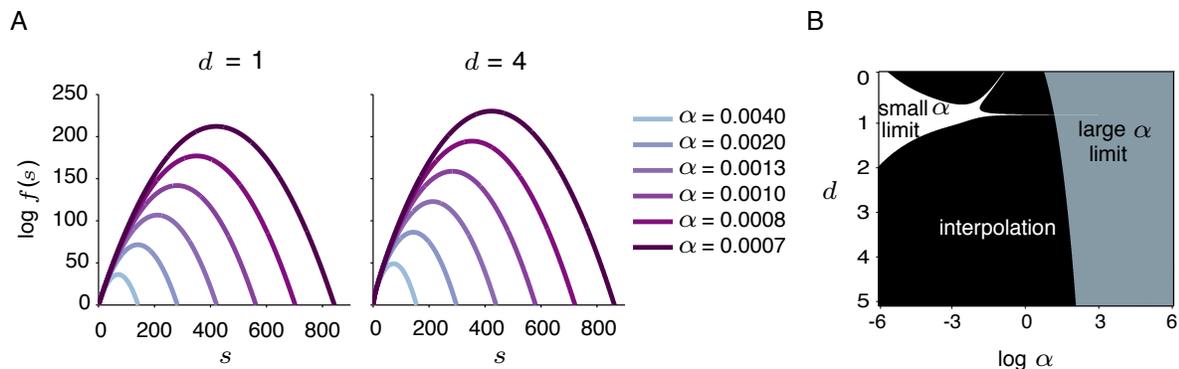
$$\log A(d, \alpha) \approx \log \int_0^\infty s^d e^{-\alpha s^2} \frac{e^s}{\sqrt{2\pi s}} ds = \frac{1}{4\alpha} - d \log \alpha - d \log 2.$$

When  $\alpha > \alpha_\infty(d)$  we approximate  $I_0(s)$  by its Taylor series around  $s = 0$  (Abramowitz and Stegun, 1965):  $I_0(s) \approx \sum_{i=0}^\infty \frac{1}{\Gamma(i+1)^2} \left(\frac{s}{2}\right)^{2i}$ , so that  $A(d, \alpha) \approx \sum_{i=0}^\infty \frac{2^{-2i}}{\Gamma(i+1)^2} \int_0^\infty s^{d+2i} e^{-\alpha s^2} ds$ .

Keeping the first two terms and evaluating the integrals, we obtain

$$\log A(d, \alpha) \approx \log 2 + \frac{d+1}{2} \log \alpha + \log \left( \Gamma \left( \frac{d+1}{2} \right) + \frac{\Gamma \left( 1 + \frac{d+1}{2} \right)}{4\alpha} \right).$$

We also build a lookup table with a million pairs of  $\alpha$  and  $d$ , and the corresponding value of  $\log A(\alpha, d)$ , which we compute numerically using Matlab's `integral` command. For  $\alpha_0(s) < \alpha < \alpha_\infty(s)$ , we evaluate  $\log A(\alpha, d)$  by linearly interpolating between entries in the table. Interpolation is a slow operation, so we replace  $I_0(s)$  with asymptotic approximations in the limit of small and large  $\alpha$ . This causes some error, which grows as  $\alpha$  deviates from 0 or  $\infty$ . We choose  $\alpha_\infty(d)$  and  $\alpha_0(d)$  such that the total error in  $\log A(\alpha, d)$  is less than 0.003 (Figure 4.17).



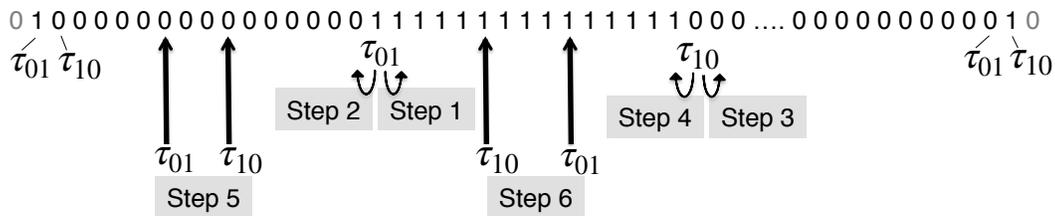
**Figure 4.17** Details of solving the integral  $A(\alpha, d) = \int f(s)ds$ , with  $f(s) = s^d e^{-\alpha s^2} I_0(s)$ .

**(A)**  $\log f(s)$  for several combinations of  $s$ ,  $d$  and  $\alpha$ . For larger values of  $\alpha$ ,  $f(s)$  is concentrated at lower values of  $s$ . For such values, we use the Taylor series expansion of  $I_0(s)$ . However, for smaller values of  $\alpha$ , the larger values of  $s$  contribute substantially to the integral and therefore we use the large  $s$  approximation of  $I_0(s)$ . These analytical approximations are much faster than interpolation, though come at the cost of approximation errors. **(B)** We limit the usage of approximations to ensure that the total approximation error of the integral  $A(\alpha, d)$  is less than 0.003. In white and gray, we show the parameter regions that satisfy this criterion.

*Markov Chain Monte Carlo sampling*

The goal of Step 3 in the BMD algorithm is to sample possible eye state time series  $\mathbf{C}$  from  $p(\mathbf{C}|\hat{\mathbf{z}}, \hat{\sigma}_x, \hat{\sigma}_z, \hat{\sigma}_0, \hat{\sigma}_1, \hat{d}_1)$ . We use an MCMC method (Newman and Barkema, 1999), which performs a biased random walk in the space of all such time series. On each step, we generate a new time series  $\mathbf{C}^{\text{new}}$ , by randomly mutating the current  $\mathbf{C}$  in one of 6 possible steps (Figure 4.18). To concisely express these steps, we reparametrized each time series  $\mathbf{C}$  in terms of its change points  $\tau$ , and separately keep track of time points where the eye state changes from drift to microsaccade ( $\tau_{01}$ ) and points where it changes from microsaccade back to drift ( $\tau_{10}$ ). The 6 steps in our MCMC sampling scheme are:

1.  $\tau_{01} \rightarrow \tau_{01} + 1$
2.  $\tau_{01} \rightarrow \tau_{01} - 1$
3.  $\tau_{10} \rightarrow \tau_{10} + 1$
4.  $\tau_{10} \rightarrow \tau_{10} - 1$
5. Create a new pair  $\tau_{01} - \tau_{10}$
6. Create a new pair  $\tau_{10} - \tau_{01}$



**Figure 4.18 Markov Chain Monte Carlo steps.** Visualization of the 6 types of steps we use to navigate the space of the eye state time series  $\mathbf{C}$ . We ensure that we take samples from the posterior probability distribution  $p(\mathbf{C}|\mathbf{x})$ .

These steps dictate the selection probability  $g(\mathbf{C} \rightarrow \mathbf{C}^{\text{new}})$ , which in general does not necessarily equal  $g(\mathbf{C}^{\text{new}} \rightarrow \mathbf{C})$ . The Markov Chain Monte Carlo algorithm accepts any of these steps with an acceptance probability  $A(\mathbf{C} \rightarrow \mathbf{C}^{\text{new}})$ . To sample from the correct posterior distribution, the Markov Chain in a Monte Carlo algorithm has to satisfy detailed balance, which ensures that the system makes transitions in and out of every state with compatible probabilities:

$$\frac{P(\mathbf{C} \rightarrow \mathbf{C}^{\text{new}})}{P(\mathbf{C}^{\text{new}} \rightarrow \mathbf{C})} = \frac{g(\mathbf{C} \rightarrow \mathbf{C}^{\text{new}}) A(\mathbf{C} \rightarrow \mathbf{C}^{\text{new}})}{g(\mathbf{C}^{\text{new}} \rightarrow \mathbf{C}) A(\mathbf{C}^{\text{new}} \rightarrow \mathbf{C})} = \frac{p(\mathbf{C}^{\text{new}} | \hat{\mathbf{z}}, \hat{\sigma}_x, \hat{\sigma}_z, \hat{\sigma}_0, \hat{\sigma}_1, \hat{d}_1)}{p(\mathbf{C} | \hat{\mathbf{z}}, \hat{\sigma}_x, \hat{\sigma}_z, \hat{\sigma}_0, \hat{\sigma}_1, \hat{d}_1)}.$$

We guarantee detailed balance using a modified Metropolis-Hastings acceptance probability (Metropolis et al., 1953; Newman and Barkema, 1999):

$$A(\mathbf{C} \rightarrow \mathbf{C}^{\text{new}}) = \min \left( 1, \frac{g(\mathbf{C}^{\text{new}} \rightarrow \mathbf{C}) p(\mathbf{C}^{\text{new}} | \hat{\mathbf{z}}, \hat{\sigma}_x, \hat{\sigma}_z, \hat{\sigma}_0, \hat{\sigma}_1, \hat{d}_1)}{g(\mathbf{C} \rightarrow \mathbf{C}^{\text{new}}) p(\mathbf{C} | \hat{\mathbf{z}}, \hat{\sigma}_x, \hat{\sigma}_z, \hat{\sigma}_0, \hat{\sigma}_1, \hat{d}_1)} \right).$$

Coarsely, this rule accepts all steps which increase the posterior probability of the new time series, and accepts some steps which decrease its posterior. However, the acceptance probability also contains a term  $\frac{g(\mathbf{C}^{\text{new}} \rightarrow \mathbf{C})}{g(\mathbf{C} \rightarrow \mathbf{C}^{\text{new}})}$  which compensates for any mismatch in selection probabilities between transitions and their reverse. This compensation term allows the Metropolis-Hastings to be flexible, and ensures detailed balance for any choice of steps.

### *Parallel tempering*

We have a high-dimensional problem with a complicated probability landscape that can be hard for Metropolis algorithms to navigate without getting stuck in local maxima. To avoid this, we performed parallel tempering (Earl and Deem, 2005; Newman and Barkema, 1999), also called replica exchange MCMC sampling), which entails perform-

ing the above Metropolis Hastings algorithm concurrently at several inverse temperatures  $\beta$ , which modify the acceptance probability to:

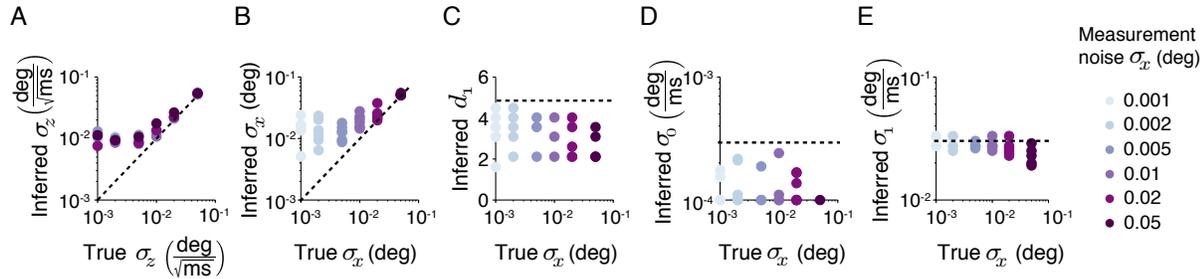
$$A(\mathbf{C} \rightarrow \mathbf{C}^{\text{new}}) = \min \left( 1, \frac{g(\mathbf{C}^{\text{new}} \rightarrow \mathbf{C}) p(\mathbf{C}^{\text{new}})}{g(\mathbf{C} \rightarrow \mathbf{C}^{\text{new}}) p(\mathbf{C})} \left( \frac{p(\hat{\mathbf{z}}|\mathbf{C}^{\text{new}}, \hat{\sigma}_x, \hat{\sigma}_z, \hat{\sigma}_0, \hat{\sigma}_1, \hat{d}_1)}{p(\hat{\mathbf{z}}|\mathbf{C}, \hat{\sigma}_x, \hat{\sigma}_z, \hat{\sigma}_0, \hat{\sigma}_1, \hat{d}_1)} \right)^\beta \right),$$

where we have split up the posterior into a prior and a likelihood. The lower the temperature (increased  $\beta$ ), the less likely it is for the Markov Chain to accept steps which reduce the likelihood. Therefore, low-temperature chains are strongly attracted by likelihood maxima (local or global), whereas high-temperature chains explore the space more widely. In the infinite-temperature limit, the Markov Chain samples from the prior  $p(\mathbf{C})$ . The strength of parallel tempering consists in allowing neighboring chains to exchange information by attempting to swap their configurations and accepting swaps with a probability

$$A(\{\mathbf{C}_1, \mathbf{C}_2\} \rightarrow \{\mathbf{C}_2, \mathbf{C}_1\}) = \min \left( 1, \left( \frac{p(\hat{\mathbf{z}}|\mathbf{C}_1, \hat{\sigma}_x, \hat{\sigma}_z, \hat{\sigma}_0, \hat{\sigma}_1, \hat{d}_1)}{p(\hat{\mathbf{z}}|\mathbf{C}_2, \hat{\sigma}_x, \hat{\sigma}_z, \hat{\sigma}_0, \hat{\sigma}_1, \hat{d}_1)} \right)^{\beta_2 - \beta_1} \right).$$

This acceptance probability ensures that we always swap if a hotter chain has stumbled on a state with a higher posterior, thus providing the algorithm with a very high chance to not get stuck in a local maxima, while ensuring that the chain with  $\beta = 1$  samples from the correct posterior probability distribution  $p(\mathbf{C}|\mathbf{z}, \sigma_z, \sigma_1, d_1, \sigma_0, d_0)$ . We choose the set of temperatures in our simulation to span the full range between  $\beta = 0$  and  $\beta = 1$ , with significant overlap in the distribution of posterior values between successive chains, so that swaps are accepted with a non-zero probability.

## 4.5.2 Parameter inference in simulated and DPI data

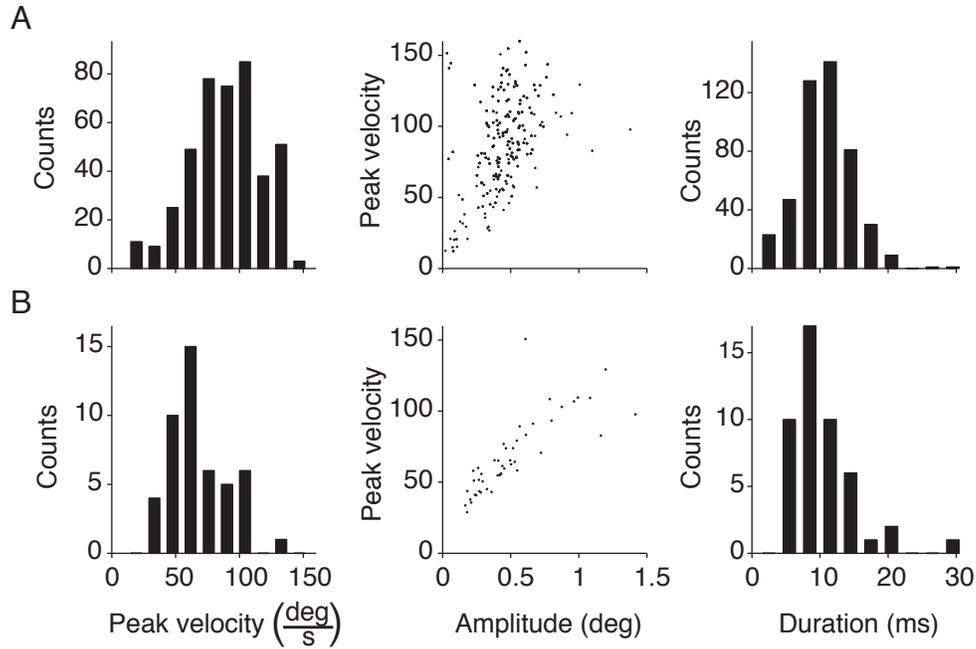


**Figure 4.19 Parameter recovery in simulated data.** In all simulated datasets, we fixed the velocity distribution parameters at  $d_1 = 4.4$ ,  $\sigma_0 = 0.0003 \frac{\text{deg}}{\text{ms}}$  and  $\sigma_1 = 0.03 \frac{\text{deg}}{\text{ms}}$ . For every combination of 6 motor noise values and 6 measurement noise values (colors), we created 8 datasets. Here we show the median across the 8 datasets of the inferred parameter values as a function of the true value of the same parameter (**A**: motor noise, **B**: measurement noise), or as a function of the true measurement noise  $\sigma_x$ , in the case of the velocity distribution parameters (**C**:  $d_1$ , **D**:  $\sigma_0$ , **E**:  $\sigma_1$ ). The dashed black lines correspond to perfect parameter recovery. While these parameters are not always faithfully recovered, the inferred eye state time series  $C$  is recovered to a great degree of accuracy (Figure 4.4).

Subject	$\hat{\sigma}_z$	$\hat{\sigma}_x$	$\hat{d}_1$	$\hat{\sigma}_0$	$\hat{\sigma}_1$
Eyelink S1	0.01397	0.0249	1.1	0.0001	0.0275
Eyelink S2	0.00723	0.0165	1.1	0.0001	0.0131
Eyelink S3	0.01317	0.10379	4.961	0.0001	0.00758
Eyelink S4	0.01265	0.0182	1.1	0.0001	0.0001
Eyelink S5	0.00637	0.02327	1.1	0.0158	0.0331
DPI S1	0.014	0.025	1.1	0.0001	0.0275
DPI S2	0.014	0.025	1.1	0.0001	0.0275

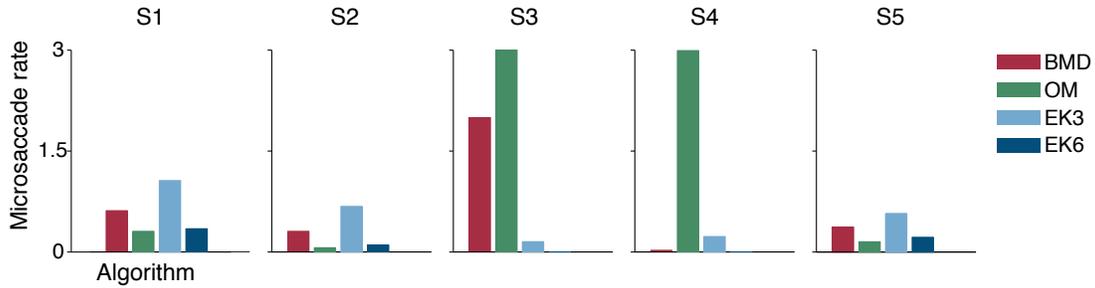
**Table 4.2 Parameter inference for EyeLink dataset 1 and DPI dataset.**

### 4.5.3 Microsaccade kinematics in the EyeLink pilot data

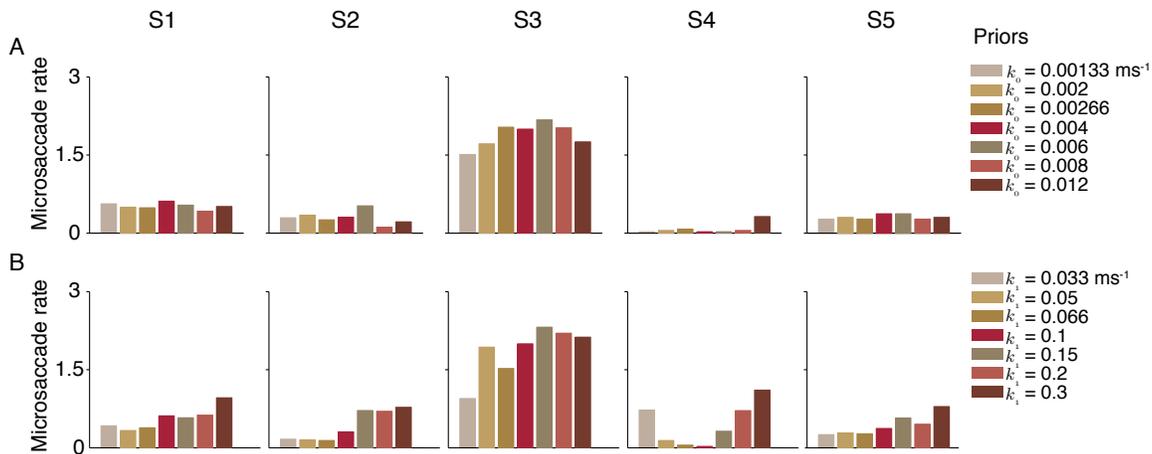


**Figure 4.20** Microsaccade kinematics in EyeLink data. Microsaccade kinematics as detected in EyeLink data with the **(A)** BMD algorithm. (*Left*) The peak velocity distributions, (*Middle*) Main sequence linear relationship between peak velocity and amplitude and (*Right*) duration distributions. **(B)** EK6 algorithm. Mostly, we notice similarities between the kinematics of the sequences detected with the two different algorithms. We spot the velocity threshold for the peak velocity distribution for the microsaccades detected by EK6.

#### 4.5.4 Microsaccades inferred in the EyeLink pilot data

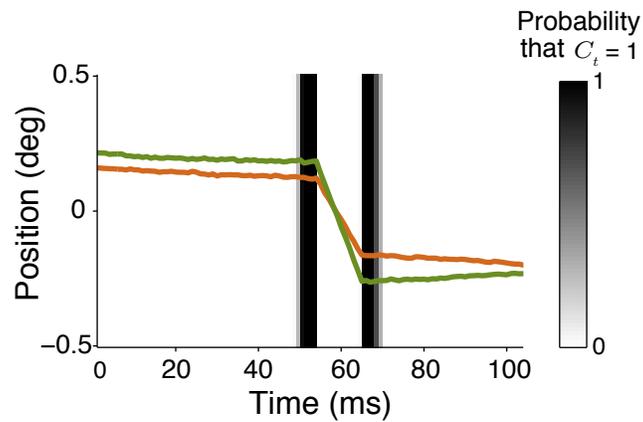


**Figure 4.21** Inferred microsaccade rates in EyeLink data vary across algorithms. Colors for the 4 algorithms are the same as in previous figures. S1 - S5 represent the five subjects.



**Figure 4.22** Inferred microsaccade rate in EyeLink data is robust to prior parameters. (A) As we vary  $k_0$ , the parameter that controls the drift duration prior, the inferred microsaccade rate varies only slightly. The lowest value  $k_0 = 0.012 \text{ ms}^{-1}$  corresponds to a drift duration distribution with median 80 ms, and the highest value  $k_0 = 0.00133 \text{ ms}^{-1}$  to 760 ms. (B) The inferred microsaccade rate does not depend too much on  $k_1$  (with the exception of subject S4). The highest and lowest values of  $k_1$  corresponds to median microsaccade durations of 3.3 ms and 30.3 ms, respectively. The somewhat larger dependence of the microsaccade rate on  $k_1$  makes intuitive sense as increasing  $k_1$  allows for very short high-velocity sequences to be labeled as microsaccades.

#### 4.5.5 BMD caveat: inference on low-noise simulated data



**Figure 4.23** Typical failure mode of BMD in low-noise simulations. Instead of detecting the microsaccade labeled by EK6, BMD detects a microsaccade right before and another microsaccade right after. This error occurs because the Kalman smoother (Step 2) converts the discontinuities at the beginning and end of the changes points into more gradual slopes, and the subsequent eye state estimation algorithm (Step 3) infers these slopes to be low-velocity microsaccades. A truly optimal inference algorithm, which marginalizes over the eye position, will not make this error.

# Chapter 5

## Discussion

In this dissertation, we investigated aspects of visual attention by utilizing new variant tasks based on classic paradigms, applying models and developing new measures. In this section, we summarize the main findings, their implications, as well as suggest future directions.

Overall, here we have applied models to dissociate processes involved in the visual memory and perceptual decision-making of both neurotypical and ADHD observers, as well as to dissociate two types of fixational eye movements, drift and microsaccades. We provide additional indication for the dissociation of the encoding and response selection stages in perceptual decision-making by their differential correlation with microsaccades. Specifically, we believe the results presented here add to the visual attention and fixational eye movements literature in the following ways: 1) show that an ideal observer model with a variable precision encoding stage and an optimal decision rule can also capture observers' data in a more naturalistic target localization task, 2) emphasize that perceptual variability during a demanding task is higher in ADHD, along with worse executive control, and can be a candidate diagnosis marker, 3) put forward a novel Bayesian microsac-

cade detection method, BMD, more robust to measurement noise from the eye tracker and 4) identify a possible mechanism for higher perceptual variability in ADHD through less effective microsaccadic suppression around stimulus onset.

In Chapter 2, we explored the effect of set size, task (detection vs localization), time (perception vs memory) and space (distant vs nearby stimuli) on visual search with heterogeneous distractors. Performance degraded with increasing target-most similar distractor (MSD) orientation distance, but not with target-MSD spatial distance. Performance did not seem to decrease with increasing distractor heterogeneity. An ideal-observer model with a variable precision encoding stage and optimal decision rule was able to capture localization data in both perception and memory. Performance decreased with the set size of the search array; mean precision decreased with the set size of the array as well. As expected, precision was higher in perception than memory. Joint fits for localization and detection data were good, suggesting that observers might be using the same encoding processes across the two tasks, as well as the respective decision rules derived from the ideal-observer models. We found a similar pattern of results with visual search arrays with reduced stimulus spacing. Observers achieved comparable performance, albeit with increased reaction times. In future analyses, we could compare the distributions of precisions across these task conditions.

In Chapter 3, we designed a demanding task that required observers to switch both covert spatial and feature attention; we created stimuli that were parametrically manipulated, thus making the data amenable to being fitted with a very simple base model, the psychometric curve. This approach, while routine in perceptual psychophysics, was only integrated in the study of ADHD in a handful of studies before us (Friedman-Hill et al., 2010; Kim et al., 2014a,b). This task design yielded two main metrics: the perceptual vari-

ability parameters from the psychometric curves, and the task-irrelevant motor output, capturing executive control. We found higher perceptual variability in ADHD, as well as higher TIMO and correlations between these metrics, suggesting shared neural sources of lower-level and higher-level deficits in ADHD. As previous studies failed to find perceptual differences in ADHD in tasks that did not also tax executive or attention processes (see (Fuermaier et al., 2017) for a review), it is possible that such differences only arise when higher-level brain processes are simultaneously engaged. Nevertheless, perceptual variability turned out to be a useful predictor for diagnosis. Based on perceptual variability alone (TIMO), we were able to classify participants into ADHD and Controls with 77% cross-validated mean accuracy, value which increased to 77.7 % when also taking into account TIMO. Thus, our simple model applied to data from a novel perceptual-executive task allowed us to quantify a difference in the perceptual encoding of ADHD participants.

Our results can be couched within broader debates in the ADHD literature of whether there is a broad diffuse deficit, or whether there is a reduced signal-to-noise ratio in higher-level brain regions. We extend the list of possible regions with reduced signal-to-noise ratio to lower brain regions that could give rise to the result of increased perceptual variability in ADHD, such as the sensory cortex, the thalamus, or even lower, the superior colliculus or locus coeruleus. This lower-level deficit could be merely a reflection of higher-level impairments; still, at a minimum, we hope to emphasize that perceptual function, at least examined while higher-level processes are simultaneously taxed, warrants further study. Our sample size was relatively small, 20 ADHD and 20 Control participants; additionally, our task design can benefit from improvements. Furthermore, usage of electroencephalography (EEG) or magnetoencephalography (MEG) recordings of neural activity while participants are engaged in a similar task can attempt to find a neural correlate for the perceptual variability parameter and furthermore, by harnessing temporal speci-

ficity and decomposing variability into early/late, (Dinstein et al., 2015; Gonen-Yaacovi et al., 2016; Mostert et al., 2015) possibly making strides towards understanding the relative contributions of sensory (early) vs decision (late) noise. To go from a promising task and candidate psychomarker to a clinically useful computational psychiatry assay several steps are needed, ultimately resembling the phases of a drug discovery pipeline (Paulus et al., 2016). For a task to successfully pass through the early stages of this pipeline, strong psychometric properties are needed, which can be checked by examining task variants, extensive checks of parameter and model identifiability and importantly test-retest reliability (Hitchcock et al., 2017).

More broadly, several tasks in the computational psychiatry of ADHD have focused on variants of the CPT task or reinforcement learning tasks. Just like a balance between exploration and exploitation is beneficial for life, this might be true for a field. Established tasks have numerous advantages: the parameters are easily interpretable, results can be relatively smoothly connected to previous work and differences can be attributed to a confined space of manipulations. It is essential to complement this work within the paradigm with exploration of novel paradigms. Low-level perception has been understudied in computational psychiatry, perhaps for good reasons as psychiatric disorders symptoms are mainly manifested in higher level cognition; however, here we showed that varying low-level stimuli parametrically can be joined with cognitive control. Applying such tasks more extensively has the additional advantage that the information flow has been (Mante et al., 2013; Siegel et al., 2015; Wimmer et al., 2015) or can be investigated neurophysiologically in animals. From a pragmatic standpoint, heterogeneity in high-level cognitive functions supported by various networks that have feedback projections into lower-level regions might be captured by low-level perceptual correlates. Thus, we hope for more extensive integration of similar psychophysical paradigms which use easily parametrizable low-level

stimuli into the study of computational psychiatry.

We also want to emphasize looking at individual differences beyond group differences, specifically collapsing across groups and looking at correlations across behavioral metrics as well as of behavioral metrics with clinical metrics. Future application of these approaches combined with formal analyses could address whether ADHD participants represent a distinctly clustered subset, or whether they fall on a continuum of behavior with controls (McLennan, 2016; Salum et al., 2014). Specifically, the two main symptom clusters in ADHD, inattention and hyperactivity/impulsivity, might be linked to different behavioral measures, perhaps one perceptual and the other more cognitive. This was not the case in our dataset (Section 3.5.6), but future studies can more thoroughly search for or design targeted tasks for behavioral correlates of each symptom cluster. In other words, this would maximize the usage of psychiatry in computational psychiatry.

Chapter 4 shows that using an inference algorithm based on a Bayesian model of microsaccade generation amongst drift could significantly improve microsaccade detection. The BMD algorithm is both more principled and produces the lowest errors on both simulated data and real data. In particular, it is substantially more robust to measurement noise (which is especially useful given the relatively high measurement noise of widely used video-based infrared eye trackers). The BMD algorithm can be extended to build in more knowledge about the processes underlying microsaccades. We have demonstrated the usefulness of BMD by applying it to the eye movement traces recorded while participants performed the task designed in Chapter 3. One caveat is that the applicability of BMD is limited to datasets with high measurement noise. Our generative model is wrong, like all models. Model mismatch, in our case the inaccuracy of the piece-wise linear assumption for the shape of the microsaccades, is not so problematic at high noise, but becomes

so when the noise is removed with some heuristic filter. This is the case for several already existent data sets that have been already filtered online, for instance with Eyelink’s heuristic filter. We are not alone in recommending collecting full noise datasets and performing filtering offline for better user control; we hope that in the future more datasets would be amenable to be analyzed with BMD.

Applying BMD to the eye movement time series recorded while participants performed the task described in Chapter 3, we find differential correlations of microsaccade rate around stimulus onset with perceptual variability, but not with TIMO, providing evidence that the task parameters capture the separable processes of perceptual encoding and stimulus-response rule selection.

Beyond our illustration of this approach in Section 4.3.5, we emphasize more generally the advantages of analyzing oculomotor parameters, here microsaccades, in clarifying the separability of perceptual encoding/processing and response selection processes in mixed behavioral tasks. Oculomotor differences in ADHD have only recently been investigated (Dankner et al., 2017; Fried et al., 2014; Munoz et al., 2003; Panagiotidi et al., 2017; Roberts et al., 2017; Wainstein et al., 2017) and not yet applied towards dissociating information processing stages in behavior. A previous illustration of the dissociative approach comes from the study of (Cavanagh et al., 2014), in which the drift rate and decision thresholds parameters in the drift diffusion model were respectively influenced by gaze dwell time and pupil dilation. While fMRI and/or EEG provide more detailed neural information than eye tracking towards dissociating the neural processes involved in behavior, they are a lot more difficult and expensive to administer at the large scales needed for computational psychiatry (Puviani et al., 2016). Thus, in the future we hope to see more task development with an eye towards dissociative processes that can be captured in ocu-

lomotor parameters applied to the study of ADHD.

# References

- (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human perception and performance*, 16(2):332–350.
- Abadi, R. V. and Gowen, E. (2004). Characteristics of saccadic intrusions. *Vision research*, 44(23):2675–2690.
- Abramowitz, M. and Stegun, I. (1965). *Handbook of Mathematical Functions*. Dover Publications.
- Acerbi, L. (2016). Bayesian adaptive stimulus placement of psychometric function for matlab.
- Acerbi, L. and Ma, W. J. (2017). Practical bayesian optimization for model fitting with bayesian adaptive direct search. *arXiv:1705.04405*.
- Adams, R. P. and McKay, D. J. (2007). Bayesian online changepoint detection. *arxiv.org/abs/0710.3742v1*.
- Adler, L. and Cohen, J. (2004). Diagnosis and evaluation of adults with attention-deficit/hyperactivity disorder. *Psychiatric Clinics of North America*, 27(2):187–201.
- Adler, L. A., Faraone, S. V., Spencer, T. J., Berglund, P., Alperin, S., and Kessler, R. C. (2017). The structure of adult ADHD. *International Journal of Methods in Psychiatric Research*, 26(1):e1555.
- Alexander, R. G., Macknik, S. L., and Martinez-Conde, S. (2018). Microsaccade characteristics in neurological and ophthalmic disease. *Frontiers in Neurology*, 9.

- Aston-Jones, G., Rajkowski, J., and Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological Psychiatry*, 46(9):1309–1320.
- Baldassi, S. and Vergheze, P. (2002). Comparing integration rules in visual search. *Journal of Vision*, 2(8):3–3.
- Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, 121(1):65–94.
- Barlow, H. B. (1952). Eye movements during fixation. *The Journal of Physiology*.
- Bashinski, H. S. and Bacharach, V. R. (1980). Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Perception & Psychophysics*, 28(3):241–248.
- Bays, P. M. (2016). Evaluating and excluding swap errors in analogue tests of working memory. *Scientific Reports*, 6(1).
- Bays, P. M., Catalao, R. F. G., and Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10):7–7.
- Benedetto, S., Pedrotti, M., and Bridgeman, B. (2011). Microsaccades and exploratory saccades in a naturalistic environment. *Journal of Eye Movement Research*.
- Betta, E. and Turatto, M. (2006). Are you ready? i can tell by looking at your microsaccades. *NeuroReport*, 17(10):1001–1004.
- Bettenbuehl, M., Paladini, C., Mergenthaler, K., Kliegl, R., Engbert, R., and Holschneider, M. (2010). Microsaccade characterization using the continuous wavelet transform and principal component analysis. *Journal of Eye Movement Research*.
- Biederman, J. (2000). Age-dependent decline of symptoms of attention deficit hyperactivity disorder: Impact of remission definition and symptom type. *American Journal of Psychiatry*, 157(5):816–818.
- Biederman J, Newcorn J, S. S. (1991). Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. *American Journal of Psychiatry*, 148(5):564–577.

- Biggs, A. T. and Mitroff, S. R. (2014). Improving the efficacy of security screening tasks: A review of visual search challenges and ways to mitigate their adverse effects. *Applied Cognitive Psychology*, 29(1):142–148.
- Bijlenga, D., Tjon-Ka-Jie, J., Schuijers, F., and Kooij, J. (2017). Atypical sensory profiles as core features of adult ADHD, irrespective of autistic symptoms. *European Psychiatry*, 43:51–57.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- Bonneh, Y. S., Adini, Y., and Polat, U. (2015). Contrast sensitivity revealed by microsaccades. *Journal of Vision*, 15(9):11.
- Boonstra, A. M., Oosterlaan, J., Sergeant, J. A., and Buitelaar, J. K. (2005). Executive functioning in adult ADHD: a meta-analytic review. *Psychological Medicine*, 35(8):1097–1108.
- Booth, J. R., Burman, D. D., Meyer, J. R., Lei, Z., Trommer, B. L., Davenport, N. D., Li, W., Parrish, T. B., Gitelman, D. R., and Mesulam, M. M. (2005). Larger deficits in brain networks for response inhibition than for visual selective attention in attention deficit hyperactivity disorder (ADHD). *Journal of Child Psychology and Psychiatry*, 46(1):94–111.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(5241):177–178.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4):433–436.
- Cameron, E., Tai, J. C., and Carrasco, M. (2002). Covert attention affects the psychometric function of contrast sensitivity. *Vision Research*, 42(8):949–967.
- Cameron, L. E., Tai, J., Eckstein, M., and Carrasco, M. (2004). Signal detection theory applied to three visual search tasks — identification, yes/no detection and localization. *Spatial Vision*, 17(4):295–325.
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51(13):1484–1525.

- Carrasco, M. (2014). *Spatial Covert Attention*. Oxford University Press.
- Casal, P. V., Esposito, F. L., Martínez, I. M., Capdevila, A., Puig, M. S., de la Osa, N., Ezpeleta, L., i Lluna, A. P., Faraone, S. V., Ramos-Quiroga, J. A., Supèr, H., and Cañete, J. (2018). Clinical validation of eye vergence as an objective marker for diagnosis of ADHD in children. *Journal of Attention Disorders*, page 108705471774993.
- Casey, B. J., Castellanos, F. X., Giedd, J. N., Marsh, W. L., Hamburger, S. D., Schubert, A. B., Vauss, Y. C., Vaituzis, A. C., Dickstein, D. P., Sarfatti, S. E., and Rapoport, J. L. (1997). Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36(3):374–383.
- Castellanos, F. X., Margulies, D. S., Kelly, C., Uddin, L. Q., Ghaffari, M., Kirsch, A., Shaw, D., Shehzad, Z., Martino, A. D., Biswal, B., Sonuga-Barke, E. J., Rotrosen, J., Adler, L. A., and Milham, M. P. (2008). Cingulate-precuneus interactions: A new locus of dysfunction in adult attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 63(3):332–337.
- Castellanos, F. X., Sonuga-Barke, E. J., Milham, M. P., and Tannock, R. (2006). Characterizing cognition in ADHD: beyond executive dysfunction. *Trends in Cognitive Sciences*, 10(3):117–123.
- Castellanos, F. X. and Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nature Reviews Neuroscience*, 3(8):617–628.
- Cavanagh, J. F., Wiecki, T. V., Kochar, A., and Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General*, 143(4):1476–1488.
- Cepeda, N. J., Cepeda, M. L., and Kramer, A. F. (2000). *Journal of Abnormal Child Psychology*, 28(3):213–226.
- Chen, C.-Y. and Hafed, Z. M. (2013). Postmicrosaccadic enhancement of slow eye movements. *Journal of Neuroscience*, 33(12):5375–5386.
- Cherici, C., Kuang, X., Poletti, M., and Rucci, M. (2012). Precision of sustained fixation in trained and untrained observers. *Journal of Vision*, 12(6):31–31.

- Ciuffreda, K. J. and Tannen, B. (1995). *Eye Movement Basics for the Clinician*. St. Louis: Mosby.
- Clements, K., Devonshire, I., Reynolds, J., and Overton, P. (2014). Enhanced visual responses in the superior colliculus in an animal model of attention-deficit hyperactivity disorder and their suppression by d-amphetamine. *Neuroscience*, 274:289–298.
- Collewijn, H. and Kowler, E. (2008). The significance of microsaccades for vision and oculomotor control. *Journal of Vision*, 8(14):20–20.
- Cornelissen, F. W., Peters, E. M., and Palmer, J. (2002). The eyelink toolbox: Eye tracking with MATLAB and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4):613–617.
- Cornsweet, T. N. (1956). Determination of the Stimuli for Involuntary Drifts and Saccadic Eye Movements. *JOSA*, 46(11):987–988.
- Crane, H. D. and Steele, C. M. (1985). Generation-v dual-purkinje-image eyetracker. *Appl. Opt.*, 24(4):527–537.
- Cubillo, A., Halari, R., Ecker, C., Giampietro, V., Taylor, E., and Rubia, K. (2010). Reduced activation and inter-regional functional connectivity of fronto-striatal networks in adults with childhood attention-deficit hyperactivity disorder (ADHD) and persisting symptoms during tasks of motor inhibition and cognitive switching. *Journal of Psychiatric Research*, 44(10):629–639.
- Dankner, Y., Shalev, L., Carrasco, M., and Yuval-Greenberg, S. (2017). Prestimulus inhibition of saccades in adults with and without attention-deficit/hyperactivity disorder as an index of temporal expectations. *Psychological Science*, 28(7):835–850.
- Daye, P. M. and Optican, L. M. (2014). Saccade detection using a particle filter. *Journal of Neuroscience Methods*, 235:157–168.
- del Campo, N., Chamberlain, S. R., Sahakian, B. J., and Robbins, T. W. (2011). The roles of dopamine and noradrenaline in the pathophysiology and treatment of attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 69(12):e145–e157.

- Devilbiss, D. M. and Berridge, C. W. (2006). Low-dose methylphenidate actions on tonic and phasic locus coeruleus discharge. *Journal of Pharmacology and Experimental Therapeutics*, 319(3):1327–1335.
- Dinstein, I., Heeger, D. J., and Behrmann, M. (2015). Neural variability: friend or foe? *Trends in Cognitive Sciences*, 19(6):322–328.
- DiStasi, L., McCamy, M. B., Catena, A., Macknik, S. L., Canas, J. J., and Martinez-Conde, S. (2013). Microsaccade and drift dynamics reflect mental fatigue. *European Journal of Neuroscience*, 38(3):2389–2398.
- Ditchburn, R. W. and Ginsborg, B. L. (1953). Involuntary eye movements during fixation. *The Journal of Physiology*, 119(1):1–17.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S. (2012). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1):27–46.
- Douglas, V. I. (1999). Cognitive control processes in attention deficit/hyperactivity disorder. In *Handbook of Disruptive Behavior Disorders*, pages 105–138. Springer US.
- Downing, C. J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance*, 14(2):188–202.
- Drugowitsch, J., Wyart, V., Devauchelle, A.-D., and Koechlin, E. (2016). Computational precision of mental inference as critical source of human choice suboptimality. *Neuron*, 92(6):1398–1411.
- Duc, A. H., Bays, P., and Husain, M. (2008). Eye movements as a probe of attention. In *Progress in Brain Research*, pages 403–411. Elsevier.
- Duncan, J. and Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3):433–458.

- Earl, D. J. and Deem, M. W. (2005). Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916.
- Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, 11(5):14–14.
- Eckstein, M. P., Thomas, J. P., Palmer, J., and Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception and Psychophysics*, 62(3):425–451.
- Engbert, R. (2006). Microsaccades: a microcosm for research on oculomotor control, attention, and visual perception. In *Visual Perception - Fundamentals of Vision: Low and Mid-Level Processes in Perception*, pages 177–192. Elsevier.
- Engbert, R. and Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision research*, 43(9):1035–1045.
- Engbert, R. and Mergenthaler, K. (2006). Microsaccades Are Triggered by Low Retinal Image Slip. *Proceedings of the National Academy of Sciences of the United States of America*, 103(18):7192–7197.
- Engbert, R., Mergenthaler, K., Sinn, P., and Pikovsky, A. (2011). An integrated model of fixational eye movements and microsaccades. *Proceedings of the National Academy of Sciences of the United States of America*, 108(39):16149–16150.
- Fair, D. A., Bathula, D., Nikolas, M. A., and Nigg, J. T. (2012). Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. *Proceedings of the National Academy of Sciences*, 109(17):6769–6774.
- Fang, Y., Gill, C., Poletti, M., and Rucci, M. (2018). Monocular microsaccades: Do they really occur? *Journal of Vision*, 18(3):18.
- Fayyad, J., Graaf, R. D., Kessler, R., Alonso, J., Angermeyer, M., Demyttenaere, K., Girolamo, G. D., Haro, J. M., Karam, E. G., Lara, C., Lépine, J.-P., Ormel, J., Posada-Villa, J., Zaslavsky, A. M., and Jin, R. (2007). Cross-national prevalence and correlates of adult attention-deficit hyperactivity disorder. *British Journal of Psychiatry*, 190(05):402–409.

- Findlay, J. M. (1971). Frequency analysis of human involuntary eye movement. *Kybernetik*.
- Fougnie, D., Suchow, J. W., and Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3:1229.
- Frank, M. J., Santamaria, A., O'Reilly, R. C., and Willcutt, E. (2006). Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neuropsychopharmacology*, 32(7):1583–1599.
- Fried, M., Tsitsiashvili, E., Bonneh, Y. S., Sterkin, A., Wygnanski-Jaffe, T., Epstein, T., and Polat, U. (2014). ADHD subjects fail to suppress eye blinks and microsaccades while anticipating visual stimuli but recover with medication. *Vision research*, 101:62–72.
- Friedman-Hill, S. R., Wagman, M. R., Gex, S. E., Pine, D. S., Leibenluft, E., and Ungerleider, L. G. (2010). What does distractibility in ADHD reveal about mechanisms for top-down attentional control? *Cognition*, 115(1):93–103.
- Fuermaier, A. B. M., Hupen, P., Vries, S. M. D., Muller, M., Kok, F. M., Koerts, J., Heutink, J., Tucha, L., Gerlach, M., and Tucha, O. (2017). Perception in attention deficit hyperactivity disorder. *ADHD Attention Deficit and Hyperactivity Disorders*.
- Fuller, S. and Carrasco, M. (2006). Exogenous attention and color perception: Performance and appearance of saturation and hue. *Vision Research*, 46(23):4032–4047.
- Geisler, W. (2004). Ideal observer analysis. In *The Visual Neurosciences*. The MIT Press.
- Geniva, L., Healey, C. G., and Enns, J. T. (2003). Target detection and localization in visual search: A dual systems perspective. *Perception & Psychophysics*, 65(5):678–694.
- Gold, J. I. and Ding, L. (2013). How mechanisms of perceptual decision-making affect the psychometric function. *Progress in Neurobiology*, 103:98–114.
- Gonen-Yaacovi, G., Arazi, A., Shahar, N., Karmon, A., Haar, S., Meiran, N., and Dinstein, I. (2016). Increased ongoing neural variability in ADHD. *Cortex*, 81:50–63.

- Gowen, E., Abadi, R., Poliakoff, E., Hansen, P., and Miall, R. (2007). Modulation of saccadic intrusions by exogenous and endogenous attention. *Brain Research*, 1141:154–167.
- Graham, N., Kramer, P., and Yager, D. (1987). Signal-detection models for multidimensional stimuli: Probability distributions and combination rules. *Journal of Mathematical Psychology*, 31(4):366–409.
- Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*.
- Greenberg, A. S., Esterman, M., Wilson, D., Serences, J. T., and Yantis, S. (2010). Control of spatial and feature-based attention in frontoparietal cortex. *Journal of Neuroscience*, 30(43):14330–14339.
- Grubb, M. A., Behrmann, M., Egan, R., Minshew, N. J., Carrasco, M., and Heeger, D. J. (2013). Endogenous spatial attention: Evidence for intact functioning in adults with autism. *Autism Research*, 6(2):108–118.
- Hafed, Z. M. and Clark, J. J. (2002). Microsaccades as an overt measure of covert attention shifts. *Vision research*, 42(22):2533–2545.
- Hafed, Z. M., Goffart, L., and Krauzlis, R. J. (2009). A neural mechanism for microsaccade generation in the primate superior colliculus. *Science*, 323(5916):940–943.
- Hafed, Z. M. and Ignashchenkova, A. (2013). On the dissociation between microsaccade rate and direction after peripheral cues: Microsaccadic inhibition revisited. *Journal of Neuroscience*, 33(41):16220–16235.
- Hafed, Z. M. and Krauzlis, R. J. (2010). Microsaccadic suppression of visual bursts in the primate superior colliculus. *Journal of Neuroscience*, 30(28):9542–9547.
- Hafed, Z. M. and Krauzlis, R. J. (2012). Similarity of superior colliculus involvement in microsaccade and saccade generation. *Journal of Neurophysiology*, 107(7):1904–1916.
- Hafed, Z. M., Lovejoy, L. P., and Krauzlis, R. J. (2011). Modulation of microsaccades in monkey during a covert visual attention task. *Journal of Neuroscience*, 31(43):15219–15230.

- Haigh, S. M., Heeger, D. J., Heller, L. M., Gupta, A., Dinstein, I., Minshew, N. J., and Behrmann, M. (2016). No difference in cross-modal attention or sensory discrimination thresholds in autism and matched controls. *Vision Research*, 121:85–94.
- Halassa, M. M. and Kastner, S. (2017). Thalamic functions in distributed cognitive control. *Nature Neuroscience*, 20(12):1669–1679.
- Halleland, H. B., Haavik, J., and Lundervold, A. J. (2012). Set-shifting in adults with ADHD. *Journal of the International Neuropsychological Society*, 18(04):728–737.
- Hauser, T. U., Fiore, V. G., Moutoussis, M., and Dolan, R. J. (2016). Computational psychiatry of ADHD: Neural gain impairments across marrian levels of analysis. *Trends in Neurosciences*, 39(2):63–73.
- Hauser, T. U., Iannaccone, R., Ball, J., Mathys, C., Brandeis, D., Walitza, S., and Brem, S. (2014). Role of the medial prefrontal cortex in impaired decision making in juvenile attention-deficit/hyperactivity disorder. *JAMA Psychiatry*, 71(10):1165.
- HenrÅquez-HenrÅquez, M. P., Billeke, P., HenrÅquez, H., Zamorano, F. J., Rothhammer, F., and Aboitiz, F. (2015). Intra-individual response variability assessed by ex-gaussian analysis may be a new endophenotype for attention-deficit/hyperactivity disorder. *Frontiers in Psychiatry*, 5.
- Herman, J. P., Bogadhi, A. R., and Krauzlis, R. J. (2015). Effects of spatial cues on color-change detection in humans. *Journal of Vision*, 15(6):3.
- Hermens, F. (2015). Dummy eye measurements of microsaccades: testing the influence of system noise and head movements on microsaccade detection in a popular video-based eye tracker. *Journal of Eye Movement Research*.
- Herrington, T. M., Masse, N. Y., Hachmeh, K. J., Smith, J. E. T., Assad, J. A., and Cook, E. P. (2009). The effect of microsaccades on the correlation between neural activity and behavior in middle temporal, ventral intraparietal, and lateral intraparietal areas. *Journal of Neuroscience*, 29(18):5793–5805.
- Hetherington, L., Dommett, E., Turner, A., Riley, T., Haensel, J., and Overton, P. (2017). Effect of methylphenidate on visual responses in the superior colliculus in the anaes-

- thetised rat: Role of cortical activation. *Journal of Psychopharmacology*, 31(10):1347–1361.
- Hitchcock, P., Radulescu, A., Niv, Y., and Sims, C. R. (2017). Translating a reinforcement learning task into a computational psychiatry assay: Challenges and strategies. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, pages 2217–2222.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van de Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- Homack, S. (2004). A meta-analysis of the sensitivity and specificity of the stroop color and word test with children. *Archives of Clinical Neuropsychology*, 19(6):725–743.
- Horowitz, T. S., Fine, E. M., Fencsik, D. E., Yurgenson, S., and Wolfe, J. M. (2007). Fixational Eye Movements Are Not an Index of Covert Attention. *Psychological Science*, 18(4):356–363.
- Huang-Pollock, C., Ratcliff, R., McKoon, G., Shapiro, Z., Weigard, A., and Galloway-Long, H. (2016). Using the diffusion model to explain cognitive deficits in attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, 45(1):57–68.
- Huang-Pollock, C. L. and Nigg, J. T. (2003). Searching for the attention deficit in attention deficit hyperactivity disorder: The case of visuospatial orienting. *Clinical Psychology Review*, 23(6):801–830.
- Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D):35–45.
- Kapoula, Z., Yang, Q., Otero-Millan, J., Xiao, S., Macknik, S. L., Lang, A., Verny, M., and Martinez-Conde, S. (2013). Distinctive features of microsaccades in alzheimer’s disease and in mild cognitive impairment. *AGE*, 36(2):535–543.

- Karalunas, S. L. and Huang-Pollock, C. L. (2013). Integrating impairments in reaction time and executive function using a diffusion model framework. *Journal of Abnormal Child Psychology*, 41(5):837–850.
- Karalunas, S. L., Huang-Pollock, C. L., and Nigg, J. T. (2012). Decomposing attention-deficit/hyperactivity disorder (ADHD)-related effects in response speed and variability. *Neuropsychology*, 26(6):684–694.
- Keshvari, S., van den Berg, R., and Ma, W. J. (2012). Probabilistic computation in human perception under variability in encoding precision. *PLoS ONE*, 7(6):e40216.
- Kessler, R. C., Adler, L., Ames, M., Delmer, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., and Walters, E. E. (2005). The World Health Organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. *Psychological Medicine*, 35(02):245–256.
- Kessler, R. C., Adler, L., Barkley, R., Biederman, J., Conners, C. K., Demler, O., Faraone, S. V., Greenhill, L. L., Howes, M. J., Secnik, K., Spencer, T., Ustun, T. B., Walters, E. E., and Zaslavsky, A. M. (2006). The prevalence and correlates of adult ADHD in the united states: Results from the national comorbidity survey replication. *American Journal of Psychiatry*, 163(4):716–723.
- Killeen, P. R., Russell, V. A., and Sergeant, J. A. (2013). A behavioral neuroenergetics theory of ADHD. *Neuroscience & Biobehavioral Reviews*, 37(4):625–657.
- Kim, S., Al-Haj, M., Chen, S., Fuller, S., Jain, U., Carrasco, M., and Tannock, R. (2014a). Colour vision in ADHD: Part 1 - testing the retinal dopaminergic hypothesis. *Behavioral and Brain Functions*, 10(1):38.
- Kim, S., Al-Haj, M., Fuller, S., Chen, S., Jain, U., Carrasco, M., and Tannock, R. (2014b). Color vision in ADHD: Part 2 - does attention influence color perception? *Behavioral and Brain Functions*, 10(1):39.
- Kim, S., Chen, S., and Tannock, R. (2014c). Visual function and color vision in adults with attention-deficit/hyperactivity disorder. *Journal of Optometry*, 7(1):22–36.

- King, J. A., Colla, M., Brass, M., Heuser, I., and von Cramon, D. Y. (2007). Inefficient cognitive control in adult ADHD: evidence from trial-by-trial Stroop test and cued task switching performance. *Behavioral and Brain Functions*, 3(1):1.
- Kingdom, F. and Prins, N. (2009). *Psychophysics: A Practical Introduction*. Academic Press, London.
- Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, 63(8):1421–1455.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., and Broussard, C. (2007). What’s new in psyctoolbox-3. *Perception*, 36(14):1.
- Ko, H., Poletti, M., and Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nature neuroscience*, 13(12):1549–1553.
- Kofler, M. J., Rapport, M. D., Sarver, D. E., Raiker, J. S., Orban, S. A., Friedman, L. M., and Kolomeyer, E. G. (2013). Reaction time variability in ADHD: A meta-analytic review of 319 studies. *Clinical Psychology Review*, 33(6):795–811.
- Kong, X., Zhen, Z., Li, X., Lu, H., Wang, R., Liu, L., He, Y., Zang, Y., and Liu, J. (2014). Individual differences in impulsivity predict head motion during magnetic resonance imaging. *PLoS ONE*, 9(8):e104989.
- Krauzlis, R. J., Lovejoy, L. P., and Zénon, A. (2013). Superior colliculus and visual spatial attention. *Annual Review of Neuroscience*, 36(1):165–182.
- Lara, A. H. and Wallis, J. D. (2012). Capacity and precision in an animal model of visual short-term memory. *Journal of Vision*, 12(3):13–13.
- Laubrock, J., Engbert, R., and Kliegl, R. (2005). Microsaccade dynamics during covert attention. *Vision Research*, 45(6):721–730.
- Laubrock, J., Engbert, R., Rolfs, M., and Kliegl, R. (2007). Microsaccades Are an Index of Covert Attention: Commentary on Horowitz, Fine, Fencsik, Yurgenson, and Wolfe (2007). *Psychological Science*, 18(4):364–366.
- Laubrock, J., Kliegl, R., Rolfs, M., and Engbert, R. (2010). When do microsaccades follow spatial attention? *Attention, Perception, & Psychophysics*, 72(3):683–694.

- Lennie, P. (2003). The cost of cortical computation. *Current Biology*, 13(6):493–497.
- Leo, D. and Gainetdinov, R. R. (2013). Transgenic mouse models for ADHD. *Cell and Tissue Research*, 354(1):259–271.
- Leth-Steensen, C., Elbaz, Z. K., and Douglas, V. I. (2000). Mean response times, variability, and skew in the responding of ADHD children: a response time distributional approach. *Acta Psychologica*, 104(2):167–190.
- Leventhal, A. G., Schmolesky, M. T., Wang, Y., and Pu, M. (2000). Degradation of stimulus selectivity of visual cortical cells in senescent rhesus monkeys. *Nature Neuroscience*, 3(4):384–390.
- Ling, S. and Carrasco, M. (2006). Sustained and transient covert attention enhance the signal via different contrast response functions. *Vision Research*, 46(8-9):1210–1220.
- Liu, T. (2003). Cortical mechanisms of feature-based attentional control. *Cerebral Cortex*, 13(12):1334–1343.
- Liu, T., Stevens, S. T., and Carrasco, M. (2007). Comparing the time course and efficacy of spatial and feature-based attention. *Vision research*, 47(1):108–113.
- Loughnane, G. M., Newman, D. P., Tamang, S., Kelly, S. P., and O’Connell, R. G. (2018). Antagonistic interactions between microsaccades and evidence accumulation processes during decision formation. *The Journal of Neuroscience*, 38(9):2163–2176.
- Lovejoy, L. P. and Krauzlis, R. J. (2017). Changes in perceptual sensitivity related to spatial cues depends on subcortical activity. *Proceedings of the National Academy of Sciences*, 114(23):6122–6126.
- Lúcio, P. S., Salum, G. A., Rohde, L. A., Swardfager, W., Gadelha, A., Vandekerckhove, J., Pan, P. M., Polanczyk, G. V., do Rosário, M. C., Jackowski, A. P., Mari, J. J., and Cogo-Moreira, H. (2016). Poor stimulus discriminability as a common neuropsychological deficit between ADHD and reading ability in young children: a moderated mediation model. *Psychological Medicine*, 47(02):255–266.
- Ma, I., van Duijvenvoorde, A., and Scheres, A. (2016). The interaction between reinforcement and inhibitory control in ADHD: A review and research guidelines. *Clinical Psychology Review*, 44:94–111.

- Ma, W. J., Husain, M., and Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3):347–356.
- Ma, W. J., Navalpakkam, V., Beck, J. M., van den Berg, R., and Pouget, A. (2011). Behavior and neural basis of near-optimal visual search. *Nature Neuroscience*, 14(6):783–790.
- Ma, W. J., Shen, S., Dziugaite, G., and van den Berg, R. (2015). Requiem for the max rule? *Vision Research*, 116:179–193.
- Ma Wei Ji, Kording Konrad, G. D. (2018). *Bayesian modeling of perception*. Textbook in preparation.
- Majdak, P., Ossyra, J. R., Ossyra, J. M., Cobert, A. J., Hofmann, G. C., Tse, S., Panozzo, B., Grogan, E. L., Sorokina, A., and Rhodes, J. S. (2016). A new mouse model of ADHD for medication development. *Scientific Reports*, 6(1).
- Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78–84.
- Martinez-Conde, S. and Macknik, S. L. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*.
- Martinez-Conde, S., Macknik, S. L., and Troncoso, X. G. (2009). Microsaccades: a neurophysiological analysis. *Trends in . . . .*
- Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., and Dyar, T. A. (2006). Microsaccades counteract visual fading during fixation. *Neuron*, 49(2):297–305.
- Martinez-Conde, S., Otero-Millan, J., and Macknik, S. (2013). The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, 14(2):83–96.
- Mazyar, H., van den Berg, R., and Ma, W. J. (2012). Does precision decrease with set size? *Journal of Vision*, 12(6):10–10.
- Mazyar, H., van den Berg, R., Seilheimer, R. L., and Ma, W. J. (2013). Independence is elusive: Set size effects on encoding precision in visual search. *Journal of Vision*, 13(5):8–8.

- McAvinue, L. P., Vangkilde, S., Johnson, K. A., Habekost, T., Kyllingsbaek, S., Bundesen, C., and Robertson, I. H. (2012). A componential analysis of visual attention in children with ADHD. *Journal of Attention Disorders*, 19(10):882–894.
- McCamy, M. B., Macknik, S. L., and Martinez-Conde, S. (2014). Different fixational eye movements mediate the prevention and the reversal of visual fading. *The Journal of Physiology*, 592(19):4381–4394.
- McLennan, J. D. (2016). Understanding attention deficit hyperactivity disorder as a continuum. *Canadian Family Physician*, 62(12):979–982.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.
- Micoulaud-Franchi, J.-A., Lopez, R., Vaillant, F., Richieri, R., El-Kaim, A., Bioulac, S., Philip, P., Boyer, L., and Lancon, C. (2015). Perceptual abnormalities related to sensory gating deficit are core symptoms in adults with ADHD. *Psychiatry Research*, 230(2):357–363.
- Mihali, A., van Opheusden, B., and Ma, W. J. (2017a). Bayesian microsaccade detection. *Journal of Vision*, 17(1):13.
- Mihali, A., Young, A. G., Adler, L. A., Halassa, M. M., and Ma, W. J. (2017b). A low-level perceptual correlate of behavioral and clinical deficits in ADHD. *bioRxiv*.
- Montague, P. R., Dolan, R. J., Friston, K. J., and Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, 16(1):72–80.
- Mostert, P., Kok, P., and de Lange, F. P. (2015). Dissociating sensory from decision processes in human perceptual decision making. *Scientific Reports*, 5(1).
- Mous, S. E., Muetzel, R. L., Marroun, H. E., Polderman, T. J. C., van der Lugt, A., Jaddoe, V. W., Hofman, A., Verhulst, F. C., Tiemeier, H., Posthuma, D., and White, T. (2014). Cortical thickness and inattention/hyperactivity symptoms in young children: a population-based study. *Psychological Medicine*, 44(15):3203–3213.

- Mueller, A., Hong, D. S., Shepard, S., and Moore, T. (2017). Linking ADHD to the neural circuitry of attention. *Trends in Cognitive Sciences*.
- Mullane, J. C. and Klein, R. M. (2008). Literature review: Visual search by children with and without ADHD. *Journal of Attention Disorders*, 12(1):44–53.
- Müller, H. J. and Findlay, J. M. (1988). The effect of visual attention of peripheral discrimination thresholds in single and multiple element displays. *Acta Psychologica*, 69(2):129–155.
- Munoz, D. P., Armstrong, I. T., Hampton, K. A., and Moore, K. D. (2003). Altered control of visual fixation and saccadic eye movements in attention-deficit hyperactivity disorder. *Journal of Neurophysiology*, 90(1):503–514.
- Munoz, D. P. and Wurtz, R. H. (1995). Saccade-related activity in monkey superior colliculus. i. characteristics of burst and buildup cells. *Journal of Neurophysiology*, 73(6):2313–2333.
- Newman, M. E. J. and Barkema, G. T. (1999). *Monte Carlo Methods in Statistical Physics*. Clarendon Press.
- Nigg, J. T. and Casey, B. J. (2005). An integrative theory of attention-deficit/ hyperactivity disorder based on the cognitive and affective neurosciences. *Development and Psychopathology*, 17(03).
- Nigg, J. T., Willcutt, E. G., Doyle, A. E., and Sonuga-Barke, E. J. (2005). Causal heterogeneity in attention-deficit/hyperactivity disorder: Do we need neuropsychologically impaired subtypes? *Biological Psychiatry*, 57(11):1224–1230.
- Nolte, L. W. and Jaarsma, D. (1967). More on the detection of one of m orthogonal signals. *The Journal of the Acoustical Society of America*, 41(2):497–505.
- Nyholt, D. R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *The American Journal of Human Genetics*, 74(4):765–769.
- Nystrom, M., Andersson, R., Niehorster, D. C., and Hooge, I. (2017). Searching for monocular microsaccades: A red hering of modern eye trackers? *Vision Research*, 140:44–54.

- Nyström, M., Hansen, D. W., Andersson, R., and Hooge, I. (2016). Why have microsaccades become larger? investigating eye deformations and detection algorithms. *Vision Research*, 118:17–24.
- Oberauer, K. and Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, 124(1):21–59.
- Ogundele, M. O., Ayyash, H. F., and Banerjee, S. (2011). Role of computerised continuous performance task tests in ADHD. *Progress in Neurology and Psychiatry*, 15(3):8–13.
- Otero-Millan, J., Castro, J., Macknik, S. L., and Martinez-Conde, S. (2014a). Unsupervised clustering method to detect microsaccades. *Journal of Vision*, 14(2):18–18.
- Otero-Millan, J., Castro, J., Macknik, S. L., and Martinez-Conde, S. (2014b). Unsupervised clustering method to detect microsaccades.
- Otero-Millan, J., Macknik, S. L., Langston, R. E., and Martinez-Conde, S. (2013). An oculomotor continuum from exploration to fixation. *Proceedings of the National Academy of Sciences*, 110(15):6175–6180.
- Otero-Millan, J., Serra, A., Leigh, R. J., Troncoso, X. G., Macknik, S. L., and Martinez-Conde, S. (2011). Distinctive features of saccadic intrusions and microsaccades in progressive supranuclear palsy. *Journal of Neuroscience*, 31(12):4379–4387.
- Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., and Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *Journal of Vision*, 8(14):21–21.
- Overton, P. G. (2008). Collicular dysfunction in attention deficit hyperactivity disorder. *Medical Hypotheses*, 70(6):1121–1127.
- Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science*, 4(4):118–123.
- Palmer, J. (2014). Divided attention. <http://faculty.washington.edu/jpalmer/files/forCM/AttentionBook/>
- Palmer, J., Ames, C. T., and Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 19(1):108–130.

- Palmer, J., Verghese, P., and Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, 40(10-12):1227–1268.
- Panagiotidi, M., Overton, P., and Stafford, T. (2017). Increased microsaccade rate in individuals with adhd traits. *Journal of Eye Movement Research*, (1).
- Panagiotidi, M., Overton, P. G., and Stafford, T. (2018). The relationship between ADHD traits and sensory sensitivity in the general population. *Comprehensive Psychiatry*, 80:179–185.
- Pastukhov, A. and Braun, J. (2010). Rare but precious: Microsaccades are highly informative about attentional allocation. *Vision Research*, 50(12):1173–1184.
- Paulus, M. P., Huys, Q. J., and Maia, T. V. (2016). A roadmap for the development of applied computational psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1(5):386–392.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10(4):437–442.
- Pestilli, F., Viera, G., and Carrasco, M. (2007). How do attention and adaptation affect contrast sensitivity? *Journal of Vision*, 7(7):9.
- Peterson, W., Birdsall, T., and Fox, W. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory*, 4(4):171–212.
- Pierre, J. M. (2013). Overdiagnosis, underdiagnosis, synthesis: A dialectic for psychiatry and the DSM. In *Making the DSM-5*, pages 105–124. Springer New York.
- Polanczyk, G., de Lima, M. S., Horta, B. L., Biederman, J., and Rohde, L. A. (2007). The worldwide prevalence of ADHD: A systematic review and metaregression analysis. *American Journal of Psychiatry*, 164(6):942–948.
- Poletti, M., Aytakin, M., and Rucci, M. (2015). Head-eye coordination at a microscopic scale. *Current Biology*, 25(24):3253–3259.
- Poletti, M., Listorti, C., and Rucci, M. (2013). Microscopic eye movements compensate for nonhomogeneous vision within the fovea. *Current Biology*, 23(17):1691–1695.

- Poletti, M. and Rucci, M. (2016). A compact field guide to the study of microsaccades: challenges and functions. *Vision research*, 118:83–97.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1):3–25.
- Puviani, L., Rama, S., and Vitetta, G. M. (2016). Computational psychiatry and psychometrics based on non-conscious stimuli input and pupil response output. *Frontiers in Psychiatry*, 7.
- Rafal, R. and Posner, M. (1987). Deficits in human visual spatial attention following thalamic lesions. *Proceedings of the National Academy of Sciences of the United States of America*, 84(20):7349–7353.
- Ratliff, F. and Riggs, L. A. (1950). Involuntary motions of the eye during monocular fixation. *Journal of experimental psychology*, 40(6):687–701.
- Rauch, H. E., Tung, F., and Striebel, C. T. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA journal*.
- Ravizza, S. M. and Carter, C. S. (2008). Shifting set about task switching: Behavioral and neural evidence for distinct forms of cognitive flexibility. *Neuropsychologia*, 46(12):2924–2935.
- Redish, A. D. and Gordon, J. A. (2016). *Computational Psychiatry: New Perspectives on Mental Illness*. MIT Press.
- Roberts, M., Ashinoff, B. K., Castellanos, F. X., and Carrasco, M. (2017). When attention is intact in adults with ADHD. *Psychonomic Bulletin & Review*.
- Robertson, C. E. and Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, 18:671–684.
- Robinson, D. L. and Wurtz, R. H. (1976). Use of an extraretinal signal by monkey superior colliculus neurons to distinguish real from self-induced stimulus movement. *Journal of Neurophysiology*, 39(4):852–870.
- Rolfs, M. (2009). Microsaccades: Small steps on a long way. *Vision Research*, 49(20):2415–2441.

- Rolfs, M., Engbert, R., and Kliegl, R. (2004). Microsaccade orientation supports attentional enhancement opposite a peripheral cue: Commentary on tse, sheinberg, and logethetis (2003). *Psychological Science*, 15(10):705–707.
- Rolfs, M., Engbert, R., and Kliegl, R. (2005). Crossmodal coupling of oculomotor control and spatial attention in vision and audition. *Experimental Brain Research*, 166(3-4):427–439.
- Rolfs, M., Kliegl, R., and Engbert, R. (2008). Toward a model of microsaccade generation: The case of microsaccadic inhibition. *Journal of Vision*, 8(11):5–5.
- Rommelse, N. N. J., Altink, M. E., de Sonneville, L. M. J., Buschgens, C. J. M., Buitelaar, J., Oosterlaan, J., and Sergeant, J. A. (2007). Are motor inhibition and cognitive flexibility dead ends in ADHD? *Journal of Abnormal Child Psychology*, 35(6):957–967.
- Rosenholtz, R. (2001). Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance*, 27(4):985–999.
- Roth, R. M., Lance, C. E., Isquith, P. K., Fischer, A. S., and Giancola, P. R. (2013). Confirmatory factor analysis of the behavior rating inventory of executive function-adult version in healthy adults and application to attention-deficit/hyperactivity disorder. *Archives of Clinical Neuropsychology*, 28(5):425–434.
- Rubia, K., Halari, R., Cubillo, A., Mohammad, A.-M., Scott, S., and Brammer, M. (2010). Disorder-specific inferior prefrontal hypofunction in boys with pure attention-deficit/hyperactivity disorder compared to boys with pure conduct disorder during cognitive flexibility. *Human Brain Mapping*, 31(12):1823–1833.
- Rucci, M., Iovin, R., Poletti, M., and Santini, F. (2007). Miniature eye movements enhance fine spatial detail. *Nature*, 447(7146):852–855.
- Rucci, M. and Victor, J. D. (2015). The unsteady eye: an information-processing stage, not a bug. *Trends in Neurosciences*, 38(4):195–206.
- Salum, G. A., Sonuga-Barke, E., Sergeant, J., Vandekerckhove, J., Gadelha, A., Moriyama, T. S., Graeff-Martins, A. S., Manfro, G. G., Polanczyk, G., and Rohde, L. A. P.

- (2014). Mechanisms underpinning inattention and hyperactivity: neurocognitive support for ADHD dimensionality. *Psychological Medicine*, 44(15):3189–3201.
- Salvucci, D. D. and Anderson, J. R. (1998). Tracing eye movement protocols with cognitive process models. *Proceedings of the twentieth annual conference of the cognitive science society*, pages 923–928.
- Salvucci, D. D. and Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on Eye tracking research and applications - ETRA*. Association for Computing Machinery (ACM).
- Santini, T., Fuhl, W., Kubler, T., and Kasneci, E. (2016). Bayesian identification of fixations, saccades, and smooth pursuits. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research and Applications - ETRA*. Association for Computing Machinery (ACM).
- Schmitt, L. I., Wimmer, R. D., Nakajima, M., Happ, M., Mofakham, S., and Halassa, M. M. (2017). Thalamic amplification of cortical connectivity sustains attentional control. *Nature*, 545(7653):219–223.
- Scholes, C., McGraw, P. V., Nyström, M., and Roach, N. W. (2015). Fixational eye movements predict visual sensitivity. *Proceedings of the Royal Society B: Biological Sciences*, 282(1817):20151568.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.
- Scolari, M., Ester, E., and Serences, J. (2014). *Feature- and Object-Based Attentional Modulation in the Human Visual System*. Oxford University Press.
- Sergeant, J. A. (2005). Modeling attention-deficit/hyperactivity disorder: A critical appraisal of the cognitive-energetic model. *Biological Psychiatry*, 57(11):1248–1255.
- Sergeant, J. A., Geurts, H., and Oosterlaan, J. (2002). How specific is a deficit of executive functioning for attention-deficit/hyperactivity disorder? *Behavioural Brain Research*, 130(1-2):3–28.

- Shen, S. and Ma, W. J. (2017). Quantifying the importance of guessing, decision noise, and variable precision in explaining behavioral variability in visual perception.
- Shin, H. and Ma, W. J. (2016). Crowdsourced single-trial probes of visual working memory for irrelevant features. *Journal of Vision*, 16(5):10.
- Siegel, M., Buschman, T. J., and Miller, E. K. (2015). Cortical information flow during flexible sensorimotor decisions. *Science*, 348(6241):1352–1355.
- Siegenthaler, E., Costela, F. M., McCamy, M. B., Stasi, L. L. D., Otero-Millan, J., Sonderegger, A., Groner, R., Macknik, S., and Martinez-Conde, S. (2013). Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *European Journal of Neuroscience*, 39(2):287–294.
- Silverstein, M. J., Faraone, S. V., Alperin, S., Leon, T. L., Biederman, J., Spencer, T. J., and Adler, L. A. (2018). Validation of the expanded versions of the adult ADHD self-report scale v1.1 symptom checklist and the adult ADHD investigator symptom rating scale. *Journal of Attention Disorders*, page 108705471875619.
- Skavenski, A., Hansen, R., Steinman, R., and Winterson, B. (1979). Quality of retinal image stabilization during small natural and artificial body rotations in man. *Vision Research*, 19(6):675–683.
- Sonuga-Barke, E. J. (2003). The dual pathway model of AD/HD: an elaboration of neurodevelopmental characteristics. *Neuroscience & Biobehavioral Reviews*, 27(7):593–604.
- Sonuga-Barke, E. J. and Castellanos, F. X. (2007). Spontaneous attentional fluctuations in impaired states and pathological conditions: A neurobiological hypothesis. *Neuroscience & Biobehavioral Reviews*, 31(7):977–986.
- Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*.
- Stampe, D. M. (1993a). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers*, 25(2):137–142.

- Stampe, D. M. (1993b). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods*.
- Stevens, A. A., Maron, L., Nigg, J. T., Cheung, D., Ester, E., and Awh, E. (2012). Increased sensitivity to perceptual interference in adults with attention deficit hyperactivity disorder. *Journal of the International Neuropsychological Society*.
- Tamber-Rosenau, B. J., Fintzi, A. R., and Marois, R. (2015). Crowding in visual working memory reveals its spatial resolution and the nature of its representations. *Psychological Science*, 26(9):1511–1521.
- Tannock, R. and Banaschewski, T. (2006). Color naming deficits and attention-deficit/hyperactivity disorder: a retinal dopaminergic hypothesis. *Behavioral and Brain Functions*.
- Terejanu, G. (2008). Crib Sheet: Linear Kalman Smoothing. Web-tutorial.
- Thome, J., Ehlis, A.-C., Fallgatter, A. J., Krauel, K., Lange, K. W., Riederer, P., Romanos, M., Taurines, R., Tucha, O., Uzbekov, M., and Gerlach, M. (2012). Biomarkers for attention-deficit/hyperactivity disorder (adhd). a consensus report of the wfsbp task force on biological markers and the world federation of adhd. *The World Journal of Biological Psychiatry*, 13(5):379–400.
- Träisk, F., Bolzani, R., and Ygge, J. (2005). A comparison between the magnetic scleral search coil and infrared reflection methods for saccadic eye movement analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 243(8):791–797.
- Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136.
- van den Berg, R., Shin, H., Chou, W.-C., George, R., and Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, 109(22):8780–8785.
- Verghese, P. (2001). Visual search and attention. *Neuron*, 31(4):523–535.
- Vincent, B. (2011). Covert visual search: Prior beliefs are optimally combined with sensory evidence. *Journal of Vision*, 11(13):25–25.

- Vincent, B. T., Baddeley, R. J., Troscianko, T., and Gilchrist, I. D. (2009). Optimal feature integration in visual search. *Journal of Vision*, 9(5):15–15.
- Visser, S. N., Danielson, M. L., Bitsko, R. H., Holbrook, J. R., Kogan, M. D., Ghandour, R. M., Perou, R., and Blumberg, S. J. (2014). Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United states, 2003–2011. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(1):34–46.e2.
- Wainstein, G., Rojas-Libano, D., Crossley, N. A., Carrasco, X., Aboitiz, F., and Ossandón, T. (2017). Pupil size tracks attentional performance in attention-deficit/hyperactivity disorder. *Scientific Reports*, 7(1).
- Welch, G. and Bishop, G. (2006). An introduction to the kalman filter. 2006. Web-tutorial.
- Wells, M. F., Wimmer, R. D., Schmitt, L. I., Feng, G., and Halassa, M. M. (2016). Thalamic reticular impairment underlies attention deficit in *ptchd1y/-* mice. *Nature*, 532(7597):58–63.
- White, A. L. and Rolfs, M. (2016). Oculomotor inhibition covaries with conscious detection. *Journal of Neurophysiology*, 116(3):1507–1521.
- Wichmann, F. A. and Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8):1293–1313.
- Wiecki, T. V., Poland, J., and Frank, M. J. (2015). Model-based cognitive neuroscience approaches to computational psychiatry. *Clinical Psychological Science*, 3(3):378–399.
- Wilens, T. E. (2004). Attention-deficit/hyperactivity disorder in adults. *JAMA*, 292(5):619.
- Wilken, P. and Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12):11.
- Willcutt, E. G. (2012). The prevalence of DSM-IV attention-deficit/hyperactivity disorder: A meta-analytic review. *Neurotherapeutics*, 9(3):490–499.

- Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., and Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry*, 57(11):1336–1346.
- Wimmer, R. D., Schmitt, L. I., Davidson, T. J., Nakajima, M., Deisseroth, K., and Hallassa, M. M. (2015). Thalamic control of sensory selection in divided attention. *Nature*, 526(7575):705–709.
- Wolfe, J. M. (2007). Guided search 4.0. In *Integrated Models of Cognitive Systems*, pages 99–119. Oxford University Press.
- Young, A. and Wimmer, R. D. (2017). Implications for the thalamic reticular nucleus in impaired attention and sleep in schizophrenia. *Schizophrenia Research*, 180:44–47.
- Yu, S.-Z. (2010). Hidden semi-markov models. *Artificial Intelligence*, 174(2):215–243.
- Yuval-Greenberg, S., Merriam, E. P., and Heeger, D. J. (2014). Spontaneous microsaccades reflect shifts in covert attention. *The Journal of Neuroscience*, 34(41):13693–13700.
- Zandbelt, B. (2014). exgauss: a matlab toolbox for fitting the ex-gaussian distribution to response time data.
- Ziegler, S., Pedersen, M. L., Mowinckel, A. M., and Biele, G. (2016). Modelling ADHD: A review of ADHD theories through their predictions for computational models of decision-making and reinforcement learning. *Neuroscience & Biobehavioral Reviews*, 71:633–656.
- Zuber, B. L., Stark, L., and Cook, G. (1965). Microsaccades and the velocity-amplitude relationship for saccadic eye movements. *Science*, 150(3702):1459–1460.