



A University of Sussex PhD thesis

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

Investigating the role of Bayesian inference in duration perception

Reny Baykova

Thesis submitted for the degree of
Doctor of Philosophy

School of Engineering and Informatics
Sussex Neuroscience 4-Year PhD Programme
University of Sussex
December 2020

Declaration

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Reny Baykova
30/12/2020

Statement

Chapter 3 and Chapter 4

Chapter 3: Section 3.1 (Model 1), and Chapter 4: overview, Section 4.1 (Introduction), and Section 4.2 (Method) have been adapted from:

Baykova, R., Buckley, C., Seth, A., & Roseboom, W. (in-principle acceptance). Influence of prior distribution variance on time perception (Registered Report Stage 1). *Royal Society Open Science*.

Authors contributions:

RB participated in the conception and design of the study, helped design the computational model, carried out the simulations, collected and analysed the data, and drafted the manuscript. CB participated in the design of the computational model and critically revised the manuscript. AS participated in the conception of the study and critically revised the manuscript. WR participated in the conception and design of the study, helped design the computational model, and helped draft the manuscript.

The manuscript for the Stage 1 registered report can be found here: <https://osf.io/75p42/>.

This work has also been presented at:

Baykova, R., Buckley, C., Seth, A. & Roseboom, W. (Aug 2019). Influence of prior distribution variance on time perception. ECVF 2019, University of Leuven, Belgium.

Baykova, R., Buckley, C., Seth, A. & Roseboom, W. (Jul 2020). How the statistical properties of stimulus distributions influence durations reproduction. FENS 2020, University of Glasgow, UK.

Chapter 5, Experiment 1

This work has been presented at:

Baykova, R. & Roseboom, W. (Sept 2019). Effect of sensory precision on behavioural and neuroimaging perceptual biases in duration estimation. CCN 2019, Technical University of Berlin, Germany. <https://doi.org/10.32470/CCN.2019.1280-0>.

Authors contributions for the conference paper (includes preliminary analysis of a subset of the data):

RB participated in the conception and design of the study, helped design the computational model, carried out the simulations, collected and analysed the data, and drafted the manuscript. WR participated in the conception and design of the study, helped design the computational model, and critically revised the manuscript.

Pre-registration can be found here: <https://osf.io/nbf38>.

Chapter 5, Experiment 2

Pre-registration can be found here: <https://osf.io/bmsc7/>.

Chapter 6

This work has been presented at:

Baykova, R., Seth, A.K. & Roseboom, W. (Jun 2018). No evidence for serial dependence in time perception outside of perceptual awareness. ASSC 22. Jagiellonian University in Kraków, Poland.

Chapter 7

Pre-registration can be found here: <https://osf.io/x9zc8>.

Data collection statement

Throughout this thesis, readers will notice that some of the reported experimental results are based on incomplete data. Due to the COVID-19 pandemic, data collection for these experiments was not possible from March to September 2020 and was resumed at a very limited capacity from September to December 2020. Where possible, results will be updated with any new data by the time of the formal examination.

Acknowledgements

I want to thank my supervisors Warrick Roseboom, Anil Seth, and Chris Buckley, for their guidance and encouragement, and for sharing their knowledge about research, academia, and life.

Thank you to my friends in Sussex and elsewhere for sticking around through the tough times.

Thank you to my fantastic partner, for his support, patience, and all the breakfasts, lunches, and dinners he prepared in the last two months.

Thank you to my family for always being there for me.

And special thanks to everyone who took part in my unbelievably long experiments – your resilience is extraordinary.

Table of Contents

Declaration.....	1
Statement	2
Acknowledgements.....	4
Thesis Summary	9
Chapter 1. General introduction	10
1.1 Perception depends on the reliability of sensory information.....	11
1.2 Perception depends on previous experience	12
1.3 Perception of time	15
1.4 Thesis overview.....	16
Chapter 2. Literature review	18
2.1 From Bayes' rule to Bayesian perceptual inference	18
2.2 Ensemble perception – forming summary representations explicitly.....	20
2.3 (Time) perception as Bayesian inference.....	23
2.3.1 Sensory likelihood – measuring time.....	23
2.3.2 Prior – building duration expectations	29
2.3.3 Posterior – integrating measured and predicted duration.....	30
2.3.4 Conclusion.....	33
2.4 Neurophysiology of time.....	34
2.4.1 Contingent negative variation	34
2.4.2 Mismatch negativity.....	36
2.4.3 Dopamine.....	39
2.4.4 Conclusion.....	42
2.5 Bayesian inference outside of awareness.....	42
2.6 Gaps in the existing literature	44
Chapter 3. Bayesian models	46
3.1 Model 1: Fixed prior, durations represented on a linear scale.....	46
3.2 Model 2: Fixed prior, durations are represented on a logarithmic scale	48
3.3 Model 3: Updating prior, durations are represented on a logarithmic scale	50
3.4 Model parameters free to vary.....	51
Chapter 4. Influence of prior distribution variance on time perception	52
4.1 Introduction	52
4.2 Method	56
4.2.1 Pre-registration and data availability.....	56
4.2.2 Design.....	56

4.2.3 Procedure.....	59
4.2.4 Materials and apparatus.....	63
4.2.5 Participant exclusion criteria	64
4.2.6 Data analysis	65
4.2.7 Required sample size	66
4.2.8 Participants	71
4.3 Results.....	71
4.3.1 Effect of variance on responses	71
4.3.2 Effect of mean on responses.....	72
4.4 Discussion.....	74
Chapter 5. Effect of sensory precision on behavioural and neuroimaging perceptual biases in duration reproduction and estimation tasks	77
5.1 Introduction	77
5.2 Experiment 1	82
5.2.1 Method	82
5.2.2 Behavioural results	94
5.2.3 EEG results	97
5.2.4 Discussion.....	101
5.3 Experiment 2	104
5.3.1 Method	105
5.3.2 Results.....	109
5.3.3 Discussion.....	111
Chapter 6. Attractive biases in duration estimation outside of awareness.....	115
6.1 Introduction	115
6.2 Methods.....	117
6.2.1 Data availability.....	117
6.2.2 Design.....	117
6.2.3 Stimuli and apparatus	118
6.2.4 Procedure.....	119
6.2.5 Participants	121
6.3 Results.....	122
6.3.1 Does duration estimation require conscious awareness?	122
6.3.2 Do durations outside of awareness contribute to regression towards the mean effects?.....	123
6.3.3 Do durations presented with and without awareness give rise to serial dependence effects?.....	124
6.4 Discussion.....	125

Chapter 7. Effect of spontaneous blinking on performance in a duration bisection task	129
7.1 Introduction	129
7.2 Method	130
7.2.1 Pre-registration and data availability.....	130
7.2.2 Design.....	130
7.2.3 Procedure.....	131
7.2.4 Apparatus.....	132
7.2.6 Analyses	133
7.2.5 Exclusion criteria	133
7.2.7 Participants	133
7.3 Results.....	134
7.4 Discussion.....	135
Chapter 8. General discussion	139
8.1 Summary of key findings.....	139
8.2 Wider implications	142
8.2.1 Reconciling attractive and repulsive perceptual biases.....	142
8.2.2 Role of sensory precision and working memory in attractive perceptual biases....	145
8.2.3 Generalizability and data quality issues in duration reproduction research: task performance	150
8.2.4 Data quality issues in duration reproduction research: equipment timing accuracy	156
8.3 Future directions.....	159
8.3.1 Do the neuronal correlates of perceptual biases in duration estimation generalise to magnitude estimation in other domains?	160
8.3.2 What are the neural correlates of the variance effect on perceptual biases in behaviour?	161
8.4 Conclusions	162
References	163
Appendix	184
Appendix A. 2IFC threshold convergence and distribution, Chapter 4	184
Appendix B. Instructions for tasks in Chapter 4.....	190
Appendix C. Testing robustness to parameter misspecification, Chapter 4.....	193
Appendix D. Instructions for tasks in Chapter 5, Experiment 1	197
Appendix E. Distribution of regression to the mean in Chapter 5, Experiments 1 and 2	199
Appendix F. 2IFC threshold convergence and distribution Chapter 5, Experiment 1	201
Appendix G. Instructions for tasks in Chapter 5, Experiment 2	208
Appendix H. 2IFC threshold distribution Chapter 5, Experiment 2	212

Appendix I. Instructions for tasks in Chapter 6.....	213
Appendix J. Instructions for tasks in Chapter 7.....	215

List of Figures

Figure 1.1. Schroeder's stairs.	13
Figure 1.2. Visualization of the results reported by Hansen et al. (2006).	14
Figure 1.3. Colour adaptation.	15
Figure 2.1. Ensemble perception of mean stimulus size across space (A), and time (B).	22
Figure 2.2. Scalar variability in time perception.	27
Figure 2.3. The two hypothesized representation of duration (or any magnitude characteristic).	29
Figure 2.4. Confounding effects of range and variance.	32
Figure 3.1. Components of Model 1.	48
Figure 4.1. Predicted effect of prior variance on perception.	55
Figure 4.2. Distributions of intervals presented in each condition.	58
Figure 4.3. Predicted effects of prior shape on regression to the mean.	58
Figure 4.4. Duration-reproduction task.	61
Figure 4.5. Behavioural results and model predictions for each experiment.	73
Figure 5.1. Regression to the mean.	79
Figure 5.2. Effect of sensory precision on perceptual biases in brain activity.	81
Figure 5.3. Stimulus conditions in the duration reproduction task.	83
Figure 5.4. Duration reproduction task.	85
Figure 5.5. 2IFC task.	87
Figure 5.6. EEG task.	88
Figure 5.7. CNV latency and electrode cluster.	93
Figure 5.8. Predictions of the Bayesian ideal observer model.	96
Figure 5.9. Effects in the behavioural data	97
Figure 5.10. Classification performance (AUC) across time, averaged over participants.	99
Figure 5.11. ERP results.	100
Figure 5.12. Effect of motor noise.	104
Figure 5.13. Correlation between sensory precision and regression to the mean.	111
Figure 6.1. Main experimental task.	122
Figure 6.2. Mean responses to the stimuli in each condition.	124
Figure 6.3. Serial dependence effects.	125
Figure 7.1. Trial structure.	132
Figure 7.2. Behavioural results.	135
Figure 7.3. Possible effect of spontaneous blinking on duration bisection.	138
Figure 8.1. Using the derivative of the Gaussian (DoG) curve to model serial dependence.	147
Figure 8.2. Perceptual biases explained by a Bayesian model (left) or DoG, (right).	149
Figure 8.3. Examples of different performance in the experiment reported in Chapter 4.	153
Figure 8.4. Distribution of device latencies.	159

List of Tables

Table 4.1. Results from simulations.	69
---	----

Thesis Summary

The brain generates predictions about the world based on our prior experiences. Such phenomena have been formally quantified through the framework of Bayesian perceptual inference. The popularity of the Bayesian framework as a theory of perception has increased greatly over the years, but there are still many questions that need to be addressed before we can ascertain whether perception can be classified as truly Bayesian. In this thesis, I investigate whether time perception follows the principles of Bayesian models of perception. The main questions I focused on are how the variability of prior expectations and individual differences in the ability to perceive durations accurately influence temporal estimation. Bayesian models suggest that the magnitude of biases towards the prior would increase if the variance of the prior decreases, but to date, this prediction has not been adequately investigated. Similarly, the theory also suggests that sensory precision, observer's ability to detect small changes in stimulus magnitude, should also affect perceptual biases, with greater sensory precision resulting in a weaker bias towards the prior. In addition, I was also interested in investigating what brain processes give rise to the perceptual biases that observers experience in magnitude estimation tasks. To do this, across different experiments, I used EEG to investigate if the brain tracks observers' subjective experience of duration, and eye-tracking to investigate the previously proposed role of dopamine in biasing duration estimation. Finally, I also investigated to what extent prior expectations and time perception, in general, are influenced by conscious awareness. Overall, the experiments presented in this thesis aim to further our understanding of how the brain constructs our perception of time and whether Bayesian frameworks constitute a useful tool for understanding perception in general.

Chapter 1. General introduction

Perception *feels* deterministic. Imagine you are at a restaurant and a waiter places a plate in front of you. You look at and smell the contents of the plate and unless you have had some weird experiences in the past, conclude that this is indeed the pizza you ordered. It looks like pizza, it smells like pizza, and you are sitting in a place that sells pizza. What else could the round doughy thing with tomato sauce, mozzarella, and pepperoni slices in front of you be? It could be many other things, for example, a very convincing pizza prop sprayed with pizza perfume. Furthermore, even if the pizza tastes precisely like a pepperoni pizza should, it might be something completely different – the dough might be gluten-free, while the pepperoni and mozzarella might be vegan. The reality is that even though perception feels deterministic, it is not. Sensory signals are ambiguous, our perceptual system is noisy, and our prior knowledge about the potential causes of the incoming sensory signals is incomplete (for example if you do not know vegan pepperoni even exists). The level of uncertainty in every step of the process dictates there cannot be a one-to-one mapping between a given sensation and its cause – the same percept could be generated by numerous processes.

The idea of probabilistic perception is not new. Helmholtz (1860) posed perception as unconscious inference – perception involves the process of inferring the most likely cause of the sensory information entering the brain. In a similar fashion, Gregory (1980) proposed that during perception, the brain engages in a process of iterative hypothesis testing in order to find the most likely interpretation of the available sensory information. It is important to note that the brain has no access to the underlying cause of sensory inputs, and this information remains hidden. Instead, the brain can only come up with its best guess regarding what process has caused this input based on previously established regularities. These ideas have been formalized mathematically through the Bayesian brain framework, which defines perception through the terms of Bayesian inference (Knill & Pouget, 2004; Pouget, Beck, Ma, & Latham, 2013). The brain forms an internal model of the world which generates predictions about the causes of incoming sensory inputs. These predictions (called priors) and the information provided by the sensory signal (the likelihood) are represented as probability density functions over the possible values the parameter of interest could take (Knill & Pouget, 2004; Pouget, Beck, Ma & Latham, 2013). The perceptual system deduces the most likely cause of sensory inputs by combining its prior predictions with the likelihood into a posterior probability distribution. Finally, the brain updates its model of the world by replacing the prior with the posterior. The extent to which the brain engages in probabilistic inference during

perception is still unknown, but extensive research has been and continues to be conducted in order to shed more light on this question.

1.1 Perception depends on the reliability of sensory information

Some of the most convincing evidence in favour of probabilistic perception comes from studies investigating multisensory integration. When multiple sources of sensory input provide conflicting information regarding the characteristics of a stimulus, the resulting percept is biased towards the source that is considered to provide more reliable information. For example, vision typically provides more accurate information about spatial location than audition (e.g., Alais & Burr, 2004). As a result, when there is a spatial discrepancy in the location information provided by the visual and auditory subcomponents of a multisensory stimulus, the visual signal usually biases the perceived location of the auditory signal – a phenomenon known as the ventriloquist effect (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003). In an extreme version of the ventriloquist effect, visual motion has even been reported to create an illusory perception of motion in an otherwise stationary auditory stimulus (Mateeff, Hohnsbein, & Noack, 2002). In contrast, audition has been proposed to provide more accurate information about temporal structure than vision (Welch & Warren, 1980), and consequently, studies have found that the timing of visual events is biased towards the timing of auditory events (Fendrich & Corballis, 2001; Vroomen & de Gelder, 2004).

Many studies have found that the integration of visuotactile and visuo-proprioceptive cues is also precision-dependent. Vision tends to provide more accurate information about object properties than touch. For example, discrimination thresholds for stimulus size were found to be lower for visual compared to haptic stimuli (Ernst & Banks, 2002). The higher precision of visual compared to tactile perception results in the dominance of visual over haptic information on reports of stimulus size (Ernst & Banks, 2002), and shape (Rock & Victor, 1964; Power & Graham, 1976; Helbig et al., 2012). Evidence consistent with probabilistic integration has been also reported in studies investigating the integration of visual and proprioceptive information to determine hand position (Van Beers, Sittig, & Gon, 1999) and movement paths (Reuschel, Drewing, Henriques, Rösler, & Fiehler, 2010), the integration of visual and vestibular cues of self-motion (Morgan, DeAngelis, Angelaki, 2008; Angelaki, Gu, & DeAngelis, 2011; see Fetsch, DeAngelis, & Angelaki, 2010 for a review), and the integration of thermal and tactile cues in reporting the temperature of an object (Ho, Chow, Tsunokake, & Roseboom, 2019).

Finally, studies have also shown that the influence of the unimodal signals on the resulting percept is dynamically adjusted when there is a change in the relative reliability of the

integrated signals. For example, as noise in the visual domain increases, the ventriloquist effect is reversed (Battaglia et al., 2003; Alais & Burr, 2004), while estimates of the size (Ernst & Banks, 2002) and shape (Heller, 1983; Helbig et al., 2012) of visuo-haptic stimuli become more biased towards the information provided in the tactile domain. These findings provide strong evidence that the brain updates its model of the world dynamically as new information is presented, which is a key element of Bayesian inference.

The literature on multisensory cue integration provides strong evidence that observers estimate the relative variability of the incoming sensory information. These results also suggest stimulus variability cannot be encoded in a modality-specific manner in the brain. For example, imagine the ventriloquist effect – you are trying to determine where in space an audio-visual stimulus originates. The auditory and visual components of the stimulus provide conflicting information and to arrive at your best guess you have to compare the reliability of the two components. If stimulus variability were encoded in a modality-specific way, it would be very difficult to make a direct comparison between the reliability of two components of the stimulus. Instead, the evidence that human observers can compare the relative variability of signals coming from different modalities suggests that the brain encodes variability in a domain-general manner.

1.2 Perception depends on previous experience

Studies on multisensory perception have shown that sensory signals are integrated optimally, taking into account their relative variability. However, perception is not influenced only by the reliability of the incoming sensory information but also by our prior knowledge about the world. The role of prior expectations in perception becomes apparent when we consider ambiguous stimuli. For example, look at Figure 1.1A. Most people will immediately interpret the image at the top as a staircase planted on the floor, facing upwards because this is how we are used to seeing staircases in the real world. However, the figure can also be interpreted as a staircase planted on the ceiling and facing downwards (Figure 1.1B). The sensory evidence in favour of both hypotheses is exactly the same, so perception cannot be based solely on the available sensory information. Instead, we use our prior knowledge about the world to help us resolve some of the existing ambiguity by shifting perception towards the interpretation with which we are more familiar.

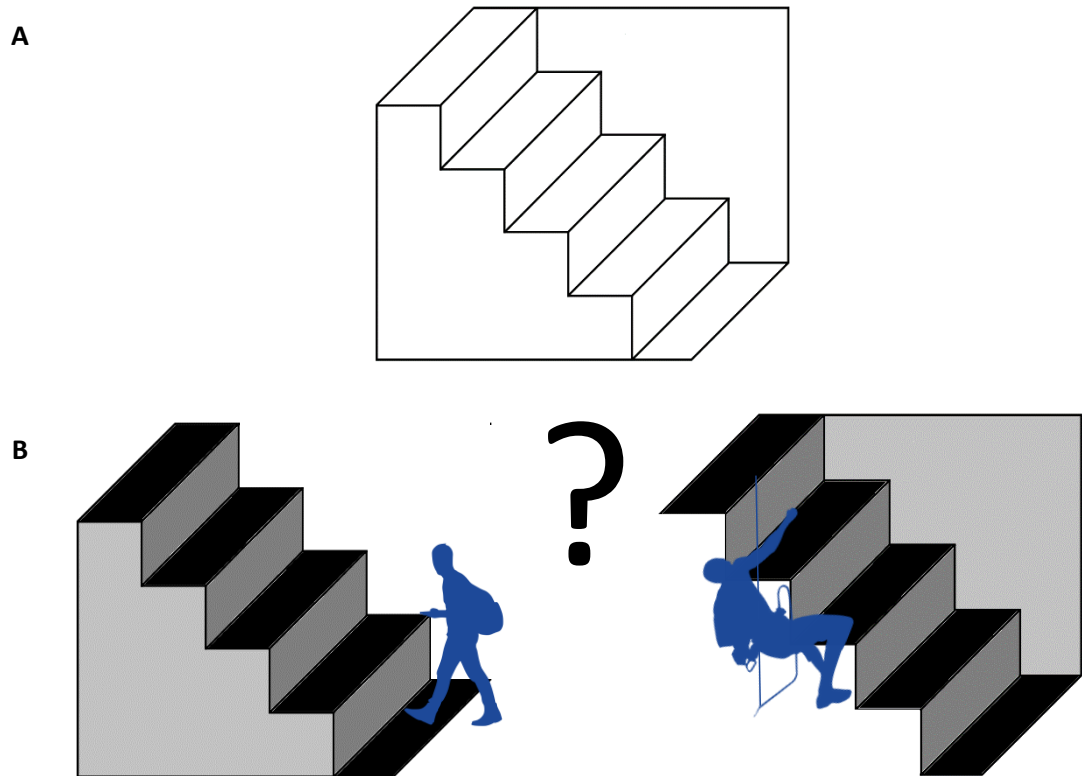


Figure 1.1. Schroeder's stairs. (A) An illusion created by Henrich Georg Friedrich Schroeder. The image is ambiguous – it can be interpreted either as normal stairs going up (bottom left image) or as stairs rotated upside-down (bottom right). **(B)** Two possible interpretations of the ambiguous image

In another relatable example of the effect of priors on perception, some everyday objects are strongly associated with a particular colour – bananas are typically yellow – which can bias observers' perception of colour. For example, when participants were asked to make images of fruit achromatic, the colour of the resulting image was shifted towards the opposite of the diagnostic colour of the fruit (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006). For the image of a banana to appear grey, the banana had to have a slightly bluish hue, while a truly achromatic image of a banana appeared slightly yellow (Figure 1.2). Studies using fMRI have further shown that observers' strong predictions about diagnostic colour are reflected in brain activity (Bannert & Bartels, 2013; Vandenbroucke, Fahrenfort, Meuwese, Scholte, & Lamme, 2014). More specifically, the diagnostic colour of objects could be decoded from activity in the visual cortex when objects were presented in grey (Bannert & Bartels, 2013), or in an ambiguous colour – for example, a tomato presented in the colour midway between red and green (Vandenbroucke et al. 2014).

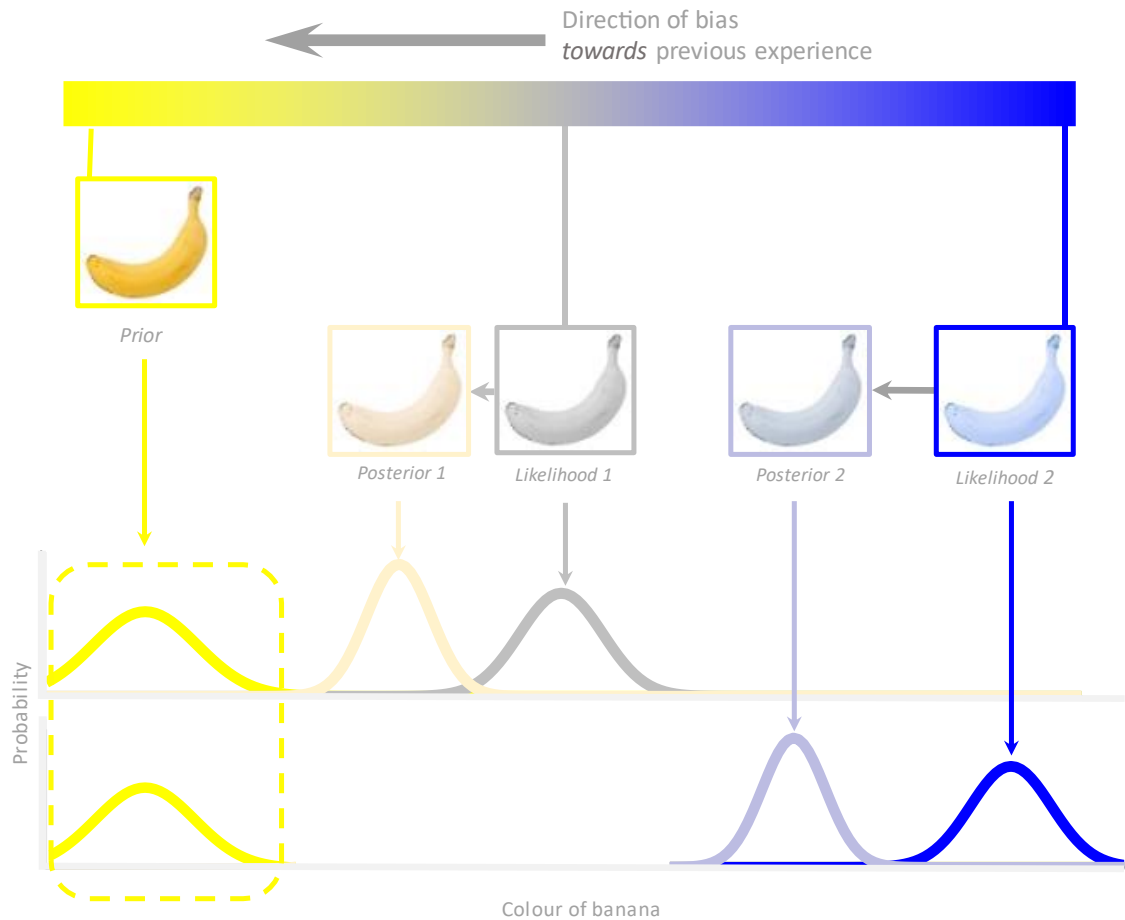


Figure 1.2. Visualization of the results reported by Hansen et al. (2006). *An achromatic image of a banana was perceived as slightly yellow, while an image with a blue tint was perceived as grey. This is just a visualization of the results and not an image generated from the analysis of the data.*

It is important to note that previous experiences can also generate repulsive biases.

Sometimes, prolonged exposure to a stimulus can bias the perception of subsequent stimuli away from the originally presented stimulus, a phenomenon known as adaptation. For example, if you stare at a yellow object for a prolonged amount of time and then look at another object, it will appear blue (Figure 1.3).

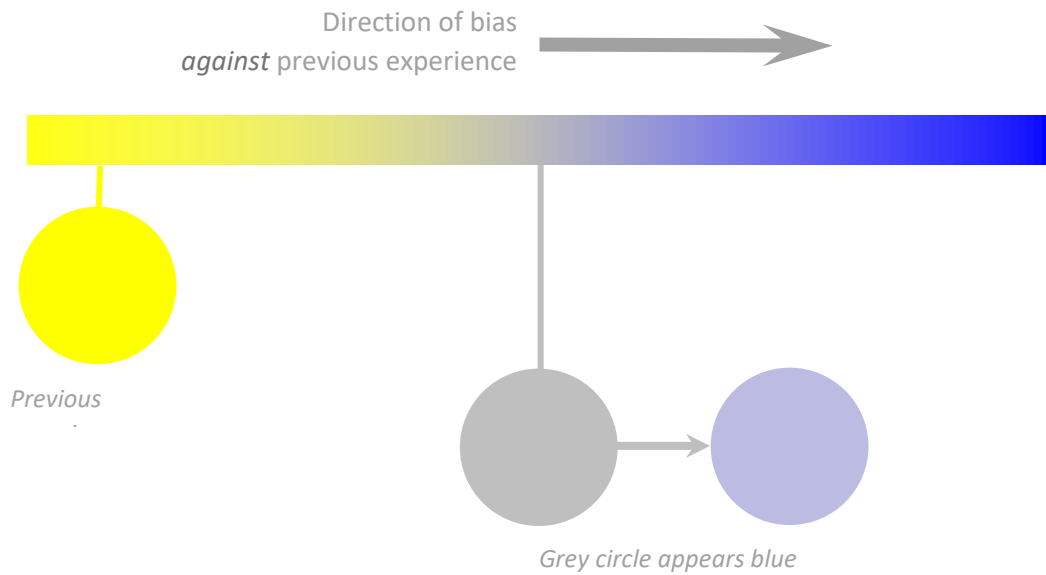


Figure 1.3. Colour adaptation. Look at the yellow circle for 20 seconds and then look at the grey circle. The grey circle should appear slightly blue.

1.3 Perception of time

In the sections above, we have discussed how sensory variability and prior expectations affect the perception of different stimulus characteristics such as colour, position, size. Would these principles also hold for more abstract stimulus features such as duration? The perception of time differs substantially from other perceptual domains in terms of its features and associated physiology. Vision, audition, and proprioception, for example, are associated with sensory-specific brain regions that are involved in the initial processing of incoming sensory-specific information. Unlike features like brightness and pitch, however, duration is not bound to any specific ‘primary’ sense – all stimuli, irrespective of sensory modality, exist in time and have temporal duration. As such, time perception has to, at least partly, be amodal and cannot be associated with one specific time-processing region the same way as basic features of vision can be associated with V1. These fundamental differences make studying the characteristics and neural underpinnings of time perception a non-trivial research endeavour.

Time is a very broad term, and it is therefore important to specify what is meant by time perception throughout this work. First, time perception operates on different scales.

Researchers have focused on three ranges – circadian timing, interval timing, and millisecond timing (Buhsu & Meck, 2005). Circadian rhythms control metabolism and sleep-wake cycles and operate over a very narrow range around 24 hours. Millisecond timing operates at a sub-second range and is involved in, for example, motor control and speech. Finally, interval timing usually refers to the perception of time of sub-second durations (although rarely below 100ms)

to minutes and even hours, and as such, it is involved in behaviours involving planning and decision-making. In addition to serving different functions, time perception across these scales also involves different neuronal mechanisms (Buhusi & Meck, 2005). The work presented here focuses on interval timing and specifically on the scale of sub- to supra-second intervals. Another important distinction in the time perception literature is the difference between retrospective and prospective timing. Usually, we do not think about the duration of a certain event until after the event has passed – for example, we do not go to the cinema with the intention to time how long the movie is, but reflect on how long the experience felt when the movie ends. This phenomenon is referred to as retrospective timing, and in such experimental tasks, participants are not informed that they will have to make a temporal judgement until the stimulus has disappeared (Brown & Stubbs, 1988). Although retrospective timing tasks have high ecological validity, they are hampered by the fact that each participant can contribute only one data point because once they have been asked to estimate the duration of a stimulus, judgements made on any following trials will no longer be retrospective. In contrast, most experiments typically use prospective timing tasks in which participants know in advance that they will be required to make temporal estimates (Brown & Stubbs, 1988). While less ecologically valid, prospective timing tasks have the advantage that participants can contribute multiple data points, and prospective estimates tend to be more accurate than retrospective estimates (Brown & Stubbs, 1988). All tasks documented here use prospective timing. In summary, here I use the term “time perception” to refer to prospective duration estimations of stimuli on the short end of the interval timing scale – milliseconds to seconds.

1.4 Thesis overview

The research discussed above shows that perception takes into account the observer’s previous experience and the relative variability of the available sensory signals. However, would the optimal integration observed across sensory signals also govern the integration of the sensory likelihood and the prior? Can manipulating the precision of the prior and the likelihood influence the magnitude of the bias exerted by the prior on perception? Also, given the fundamental differences between time perception and the primary perceptual domains, it is still not clear to what extent the perception of time is governed by these processes, and how previous experiences influence our perception of duration.

The aim of this thesis is to investigate the role of Bayesian inference in duration perception and the role that the statistical features of stimulus distributions play in the integration of priors and likelihoods. Chapter 2 provides a more detailed look at the mathematical basis of

Bayesian inference. I also review the existing evidence that observers can extract the statistical features of sensory information (range, central tendency, variance), as well as theoretical and empirical work that gives an indication of how duration estimation can be situated within a Bayesian framework. In Chapter 3, I introduce the different variations of Bayesian models used throughout the following empirical chapters. Then, in Chapter 4, I present two experiments that focused on how the shape of the prior, more specifically, the variance of the prior, affects duration reports. Our main hypothesis, based on the prediction of a Bayesian ideal observer model, was that responses would be biased towards the prior more when the prior has a smaller variance. We found that the magnitude of perceptual biases was greater in a peaked Gaussian compared to a uniform distribution, supporting the predictions of the model. However, we did not find evidence of a stronger bias in a skewed peaked distribution compared to a uniform distribution. Having established how changes in the shape of the prior influence duration reports, in Chapter 5 I present two duration perception experiments in which we investigated the effect of sensory precision on biases towards the prior in behavioural reports and brain activity, measured using EEG. In the first experiment, we found that the prior had a clear effect on both duration responses and brain activity – the same physical stimulus elicited different duration estimates and different patterns of EEG activation depending on the range of the underlying stimulus distribution. However, across both experiments, we found no relationship between sensory precision and the magnitude of either the behavioural or neuroimaging estimates of perceptual biases. All the experiments introduced so far investigated duration estimation biases induced by stimuli that participants were aware of, but it is still not known to what extent conscious awareness is necessary for the perception of durations or for the perceptual biases to occur. In Chapter 6, we investigated how masking visual stimuli affects participants' estimates of duration and whether masked stimuli give rise to perceptual biases. We found that participants were not able to estimate the duration of masked stimuli, masked stimuli did not influence the processing of subsequently presented durations, and masked stimuli were not taken into account when estimating global distributional statistics. The final empirical chapter, Chapter 7, presents our preliminary findings from a study investigating the role of dopamine in the perception of time. Finally, Chapter 8 discusses the findings reported in the previous chapters in relevance to the existing literature, highlights some wider issues in the time perception research, and provides ideas for future studies that build onto our findings.

Chapter 2. Literature review

Here I will discuss results of previous studies investigating a range of perceptual domains, but with a specific focus on empirical and theoretical work in the field of time perception. I will start this review with an overview of Bayesian statistical inference to get a better understanding of the principles behind the Bayesian brain hypothesis. I will then present research on ensemble perception to show that observers can estimate the statistical properties of stimulus distributions, and discuss how the framework of Bayesian perceptual inference can be used to understand the processes involved in duration estimation. In the end, I will discuss the behavioural and neuroimaging evidence investigating if human observers can form representations of summary statistics (and consequently engage in Bayesian inference) in the absence of awareness.

2.1 From Bayes' rule to Bayesian perceptual inference

The process of Bayesian inference is based on Bayes' rule, which is derived from the formula for calculating conditional probabilities – the probability of some event A occurring, given that another event B has occurred:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1)$$

where $P(A|B)$ is the conditional probability of A given B, $P(A \cap B)$ is the joint probability of A and B occurring together, and $P(B)$ is the unconditional probability of B occurring independently of A. The key to deriving Bayes' rule is that the joint probability $P(A \cap B)$ can be rewritten in two ways:

$$P(A \cap B) = P(A)P(B|A) \text{ but also } P(A \cap B) = P(B)P(A|B) \quad (2)$$

Therefore:

$$P(B)P(A|B) = P(A)P(B|A) \quad (3)$$

Rearranging this equation gives us Bayes' rule:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (4)$$

Within the Bayesian framework, $P(A|B)$ is called the posterior, $P(A)$ – the prior, $P(B|A)$ – the likelihood, and $P(B)$ – the marginal likelihood or evidence. To explain the intuition behind Bayesian inference, let us take an example. Imagine that you have bought a bag of almonds

from a new store, and you want to estimate the quality of the almonds given the number of bitter almonds in the bag. The bag contains 20 almonds, and the number of bitter almonds can vary from 0 (highest quality) to 20 (worst quality). As the store is new, you have no previous knowledge of the quality of almonds sold there, so your prior will take the shape of a uniform distribution – you will assign an equal probability to all possible values of quality. By eating the bag of almonds, you discover that 3 of the almonds were bitter, so the marginal likelihood is the probability of having 3 bitter almonds – $3/20$. You use this information to build the likelihood probability distribution (in this case a Binomial distribution) which gives you the probability of finding 3 bitter almonds under all possible values of quality. Having gathered this new information, you can update your initial belief about the quality of the almonds by multiplying the prior with the likelihood, which gives you the posterior probability distribution. As the marginal likelihood is just a constant, we can drop the term from the equation above, and rewrite Bayes' rule to state that the posterior is proportional to the product of the prior and the likelihood:

$$P(A|B) \propto P(A)P(B|A) \quad (5)$$

Importantly, the influence of the prior and likelihood on the posterior is weighted based on their relative precision – in this example, the likelihood has a smaller variability than the prior, so the posterior will be shifted more towards the likelihood. As our variable of interest, in this case, is discrete, the posterior probability distribution gives us the exact probability that quality is equal to each of the possible parameter values. You can therefore estimate almond quality by identifying the parameter value that has the highest probability under the posterior distribution. Finally, the process of Bayesian inference can also be iterative – your current posterior immediately becomes your prior the next time you observe new information.

If perception involves a process similar to Bayesian inference, observers must be able to track, at least roughly, the probability of different stimulus values occurring in order to build and update their prior beliefs over time. Unlike the example with almond quality above where the variable of interest is discrete, the sensory features the brain is trying to estimate are (mostly) continuous – for example distance, brightness, volume. Therefore, in the Bayesian brain framework, the prior and likelihood of observed data are represented as probability density rather than probability mass functions (e.g., Knill & Pouget, 2004). In a probability density function, the probability of a parameter having a specific value is zero. Therefore, following the Bayesian brain hypothesis, the brain cannot be tracking the specific probability of each level of, for example, brightness. It is still not clear how exactly prior expectations are represented in

the brain, but one efficient approach would be to summarize the regularities of sensory input by tracking the statistical characteristics of sensory information over time (Khayat & Hochstein, 2018) – for example, what stimulus appears most often, or how variable the sensory inputs are. By extracting the statistical properties of presented stimuli, such as their range, central tendency, and variability, the brain would be able to form an overall representation of the stimulus distribution. If we subscribe to the assumption of Gaussianity, the mean and variance of the posterior can be easily derived based on the mean and variance of the likelihood and prior:

$$\mu_{Posterior} = \frac{\mu_{Prior} \times \sigma_{Prior}^2 + \mu_{Likelihood} \times \sigma_{Likelihood}^2}{\sigma_{Prior}^2 + \sigma_{Likelihood}^2} \quad (6)$$

$$\sigma_{Posterior}^2 = \frac{\sigma_{Prior}^2 \times \sigma_{Likelihood}^2}{\sigma_{Prior}^2 + \sigma_{Likelihood}^2} \quad (7)$$

Therefore, an important prerequisite for Bayesian-like processes to take place in perception is for observers to be able to compute summary representations of stimulus features presented across time and space. Evidence that this is the case comes from decades of studies on ensemble perception.

2.2 Ensemble perception – forming summary representations explicitly

Extracting the statistical characteristics of multiple stimuli presented in a single instance in time or in a sequence across time can be useful to generate a gist of the environment. Studies on ensemble perception have shown that when asked explicitly to do so, participants can estimate the central tendency and variability of stimulus sets (for a review see Whitney & Leib, 2018).

Studies on ensemble perception provide overwhelming evidence that human observers form summary representations of central tendency when presented with multiple stimuli at the same time (Figure 2.1A). For example, when presented with a display of similar stimuli such as lines with different orientations observers can accurately judge the average orientation of the set (Dakin & Watt, 1997; Parkes et al., 2001). An abundance of studies suggest that people can perform such operations on other stimulus features ranging in complexity including direction and speed of motion (Williams & Sekuler, 1984), spatial position (Morgan & Glennerster, 1991; Hess & Holliday, 1992; Whitaker, McGraw, Pacey, & Barrett, 1996; Alvarez & Oliva, 2008), size (Ariely, 2001; Chong & Triesman, 2003, 2005; Whiting & Oriet, 2011), numerical magnitude

(Morris & Masnick, 2015), facial expression and gender (Haberman & Whitney, 2007, 2009a). Despite the evidence that observers form estimates of central tendency when presented with multiple stimuli, it is still not clear what processes observers engage in to arrive at such estimates. The experiments presented above assume that participants average across all presented stimuli (e.g. Chong & Triesman, 2003). However, evidence suggests that the mean size of an array can be accurately estimated by subsampling one to three stimuli from the set, and comparing the average stimulus size across two sets can be achieved by considering only the smallest or largest stimuli in the two sets without actually computing the mean (Myczek & Simons, 2008; although see Chong, Joo, Emmanouil & Triesman, 2008 for a response). Therefore, while there is an abundance of evidence that human observers form (some sort of) estimates of central tendency, it is not clear how these estimates are computed, and future studies should include better controls to ensure tasks investigating mean estimation cannot be completed using simpler strategies.

Research has also shown that people can extract the mean stimulus value when stimuli are presented in a sequence (Figure 2.1B). Observers can estimate the average size (Albrecht & Scholl, 2010; Corbett & Oriet, 2011; Albrecht, Scholl, & Chun, 2012; Khayat & Hochstein, 2018), brightness (Khayat & Hochstein, 2018), orientation (Khayat & Hochstein, 2018), location (Juni, Gureckis, & Maloney, 2012), facial expression (Haberman & Whitney, 2009b), numerical magnitude (Malmi & Samson, 1983), pitch (Albrecht et al., 2012; Piazza, Sweeny, Wessel, Silber, & Whitney, 2013), loudness (Curtis & Mullin, 1975), heaviness (Anderson, 1967), and duration (Schweickert, Han, Yamaguchi, & Fortin, 2014; Ren, Allenmark, Muller, & Shi, 2020) when stimuli are presented sequentially. Like averaging across space, averaging across time is also likely to be achieved without necessarily computing the mean of all stimuli, although the strategies employed are likely to be different. For example, Hubert-Wallander and Boynton (2015) found evidence that the subset of stimuli in a sequence that observers used to estimate central tendency were selected based on their position in the sequence. Estimates of average size, motion direction, and facial expression were influenced more by items presented later in the sequence, displaying recency effects (also see Anderson, 1967 for similar results regarding mean heaviness estimation). In contrast, average location estimates were influenced more by stimuli presented earlier in the sequence, displaying primacy effects. In addition to these biases, studies on averaging stimulus magnitude across time have also suggested that observers' estimates of central tendency are closer to the geometric compared to the arithmetic mean of the presented stimuli (Ren et al., 2020). Overall, very few studies have attempted to investigate what measure of central tendency participants compute when asked

to estimate the “mean” of a group of stimuli (see Watamaniuk & Duchon, 1992 for an attempt to compare the arithmetic mean and mode), and it is still not clear how these estimates are computed and what statistic they represent.

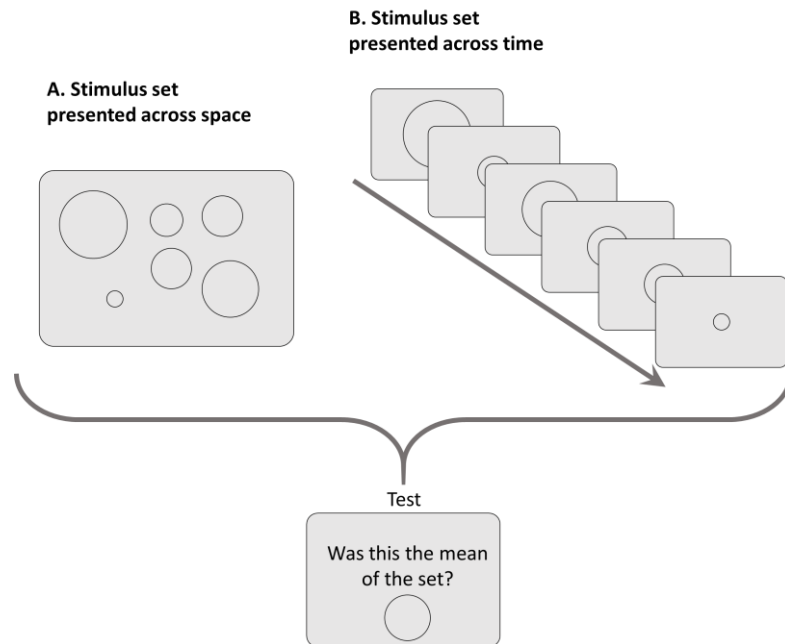


Figure 2.1. Ensemble perception of mean stimulus size across space (A), and time (B). After the presentation of the set, participants are asked to make some judgement, such as if a test probe is equal to the mean of the just presented set.

While a lot of research on ensemble perception has focused on observers’ ability to estimate central tendency, variance estimation has not been researched as extensively. It has been found that participants can estimate the variance of the facial expressions in a set of faces (Haberman, Lee, & Whiney, 2015), the orientation in a set of lines or gratings (Dakin & Watt, 1997; Morgan, Chubb, & Solomon, 2008), the size in a set of circles (Solomon, Morgan & Chubb, 2011; Tokita, Ueda, & Ishiguchi, 2016), the direction of motion in an RDK (Suárez-Pinilla, Seth, & Roseboom, 2018). In addition, long exposure to a set of stimuli also creates variance adaption – after observing an ensemble of object highly variable in colour, subsequently presented ensembles appear less variable and vice versa (Norman, Heywood, & Kentridge, 2015; Maule & Franklin, 2020). This aftereffect was further found to translate across object features from colour to orientation (Maule & Franklin, 2020), which suggests that the stimulus variance could be an abstract property encoded independently of the stimulus feature in question. To our knowledge, however, no study to date has investigated variance estimation in sensory domains outside of vision or in ensembles presented across time rather than space, so much is yet unknown about the generalizability of this process. Studies have

also shown that the variance of stimulus sets influences performance on perceptual tasks and the estimation of other ensemble statistics, specifically central tendency. For example, participants were more accurate in estimating the average direction of motion in an RDK and the average colour or shape of objects in a display when it has a smaller variability (de Gardelle & Summerfield, 2011; de Gardelle & Mamassian, 2015). In addition, Juni et al. (2012) found that estimates of the mean location of targets presented in a sequence depended on the known variance of the distributions from which each target is sampled – targets that participants knew came from distributions with smaller variance were weighted more heavily in the estimation of mean location. Together these findings suggest that one function of variance estimation could be to judge the reliability of sensory inputs to estimate other statistical properties of stimulus ensembles in an optimal manner.

Even though it is still unclear what processes participants engage in to estimate the summary statistics of stimulus distributions, the studies discussed above provide evidence that observers can estimate the central tendency and variance of stimulus sets. This suggests that the perceptual system is equipped with the necessary mechanics to engage in Bayesian perceptual inference. In the next section, I will discuss how duration estimation can emerge from a process like Bayesian perceptual inference.

2.3 (Time) perception as Bayesian inference

If we pose time perception as a process governed by the principles of Bayesian inference, we have to consider how the brain measures time, how prior expectations of duration are represented in the brain and integrated with the sensory likelihood to form perception, and finally, what signals in the brain might reflect these operations.

2.3.1 Sensory likelihood – measuring time

To build the sensory likelihood Gaussian distribution, we need to determine the mean and the variance of the distribution. The mean will correspond to the brain's estimate of the presented duration, and the variance will correspond to the noise associated with making this estimate. Here I briefly review some of the theories that have been proposed to explain how time is measured in the brain. However, it is important to note that investigating the specific process through which the brain measures time is not the core aim of the research I present in the subsequent chapters. My focus is specifically on the integration of the prior and the likelihood, for which the process underlying the likelihood is of little direct relevance.

Mean of the sensory likelihood

Multiple theories have been proposed to explain how the brain forms estimates of duration and they can be loosely divided into theories suggesting that the brain tracks time like a clock and theories showing that time can be encoded without such specialized machinery. The most prominent class of models in time perception, which can be collectively labelled under the umbrella term of scalar timing theory, propose that the brain estimates durations through an internal clock mechanism (e.g., Treisman, 1963; Gibbon, 1977). At all times, a pacemaker generates countable pulses at regular intervals (also see Miall, 1989 for an account based on multiple pacemakers). When an event of interest occurs, a switch is turned on, and an accumulator starts counting the number of pulses that have occurred since turning the switch. When the event of interest ends, a readout of the accumulator is taken and stored in memory allowing the observer to transform the accumulators' estimate into reportable units of time, compare the current accumulator estimate to previous estimates still in memory, and (if required) engage in some decision processes. Numerous other theories, however, challenge the notion that an internal clock mechanism is needed to encode time. Such theories include the state-dependent network model – durations are encoded in the state of cortical neural networks (Karmarkar & Buonomano, 2007), the firing rate model – durations are encoded in the firing rate of neurons (Durstewitz, 2003), and the duration channels model which proposes that duration is encoded in the brain through neuronal selectivity which is the same mechanisms used to encode other stimulus features like orientation and pitch (Heron, Aaen-Stockdale, Hotchkiss, Roach, McGraw, & Whitaker, 2012). Neurons in the visual and auditory cortex respond selectively to a narrow range of orientations and frequencies respectively, and knowing which neurons were activated allows the rest of the brain to infer what sensory information was presented (Hubel & Wiesel, 1968; Regan & Tansley, 1979). As a result, Heron et al. (2012) proposed that time perception is based on a similar channel-based mechanism with different populations of neurons responding selectively to a specific range of durations, and provide evidence that such a model can explain duration adaptation effects.

In addition to an internal clock being unnecessary to track time, another issue with scalar timing theory is that it assumes that the brain's internal clock mechanism attempts to track objective time. However, other work suggests that duration estimates are based on perceptual content. The estimated duration of an interval increases with increase in the number of events or the complexity of events occurring during the interval (Frankenhaeuser 1959; Ornstein, 1969; Block, 1974; Avant, Lyman, & Antes, 1975). Proponents of scalar timing theory have suggested that some mechanism based around arousal and attention that could account for

these findings. The presentation of an external stimulus leads to an increase in arousal which in turn speeds up the internal clock and results in duration overestimation (Penton-Voak, Edwards, Percival, & Wearden, 1996; Droit-Volet & Wearden, 2002). Following this proposition, the more complex the event to be estimated is, the more the speed of the clock will increase, leading to greater overestimation. While arousal controls the speed of the pacemaker, attention has been proposed to affect how quickly the switch that connects the pacemaker and accumulator is turned on (Lejeune, 1998). The accumulator starts counting the ticks emitted by the pacemaker only after the switch is turned on, so the quicker the switch is turned on, the more accurate the estimate by the accumulator will be. It has been proposed that when attention is low, the latency between the start of the event to be timed and the turning of the switch increases and becomes more variable, leading to underestimation of duration and more variable estimates respectively (Droit-Volet, 2003; Allan, 1992). Others have proposed the addition of another component to the internal clock model, an attentional gate situated between the pacemaker and the switch which opens more or less depending on how much attention is paid to the duration of an event (Zakay & Block, 1996). When the observer pays attention to the duration of the stimulus, the gate opens more widely and allows more pulses to pass through to the accumulator. There is still no consensus amongst proponents of scalar timing whether attention modulates time estimates through a separate gate component or by controlling the switch in the original formulation of the internal clock model (Zakay, 2000; Lejeune, 2000).

Even though mechanisms based around arousal and attention could help scalar timing theory account for the discrepancy between objective and subjective time, a more parsimonious explanation for this deviation has been proposed by a new model of time perception suggesting that the brain estimates time by directly keeping track of changes in perceptual content without the involvement of any clocks (Roseboom et al., 2019; Sherman, Fountas, Seth, & Roseboom., 2020). The researchers used a hierarchical deep convolutional classification network to develop a model that estimates the duration of visual scenes. Changes in perceptual content drive changes in the activation patterns of the units throughout the networks. If the change in activation exceeds some threshold, time is perceived to have passed. At the end of the event of interest, the accumulated units of time are transformed into reportable units such as milliseconds or seconds. To test the explanatory power of this theory, the researchers presented human observers and the neural network with the same naturalistic videos and found that the temporal estimates generated by the model corresponded well to the estimates reported by the participants. In conclusion, the proposition

that time is tracked based on the content of perceptual experience is more parsimonious than the time-sensing machinery proposed by scalar timing theory, and this study provides further evidence that the brain does not need an internal clock to track time. However, which mechanism the brain employs to measure time (counting ticks emitted by a pacemaker or counting changes in the environment) has no implications for the purposes of investigating how prior expectations influence perceived duration.

Standard deviation of the sensory likelihood

After we have determined the different possible ways in which the brain could be estimating time, we have to discuss what processes determine the reliability of time estimates. As studies on multisensory integration suggest (see *General Introduction*, p.11), the variance of the sensory likelihood will partially depend on how noisy the stimulus is, with noisier stimuli leading to broader sensory likelihoods (e.g., Battaglia et al., 2003; Alais & Burr, 2004). Assuming that the noise in the sensory signal is kept constant, the variance of the sensory likelihood will depend predominantly on the resolution with which the perceptual system can make the required estimate. Here I will call this resolution sensory precision – the greater the sensory precision, the more accurate the judgements produced by the perceptual system will be, and the lower the variance of the sensory likelihood. Take a clock face as an example. A clock-face with markings showing the hours and minutes will allow you to differentiate smaller differences in time than a clock face with no markings. The former clock face will therefore allow you to tell what time it is with greater precision, and the probability distribution of the current time will have a smaller variance.

In perceptual tasks, including duration estimation, sensory precision can be inferred following Weber's law. Weber's law postulates that the just noticeable difference between two stimuli scales with stimulus magnitude – the incremental increase in stimulus magnitude required for an observer to notice a change in the stimulus is proportional to the absolute magnitude of the stimulus. Specifically, Weber's law states that:

$$k = \frac{\Delta I}{I} \quad (8)$$

where the constant k , known as the Weber fraction, describes the proportional relationship between the absolute stimulus magnitude I and the just noticeable difference ΔI associated with I . In other words, the smaller the Weber fraction, the greater the sensory precision associated with a given stimulus. It is assumed that the Weber fraction is different across tasks and modalities, for example, distance discrimination versus contrast discrimination, but within

a given sensory task it is believed to be independent of absolute stimulus magnitude. Within the time perception literature, Weber's law is also widely referred to as the scalar property of variance (scalar variability for short). Scalar variability refers specifically to the observed phenomenon that as stimulus' duration (or any magnitude-related property) increases, responses become more variable (Figure 2.2; Gibbon, 1977). This relationship has been quantified through the coefficient of variation which represents the ratio between the standard deviation of the produced estimates and either the duration of the interval being estimated or the mean of the provided responses (e.g., Gibbon, 1977; Wearden, 1991). As the interval to be estimated increases, so should the associated variance of estimates, such that the coefficient of variation remains constant across a range of durations.

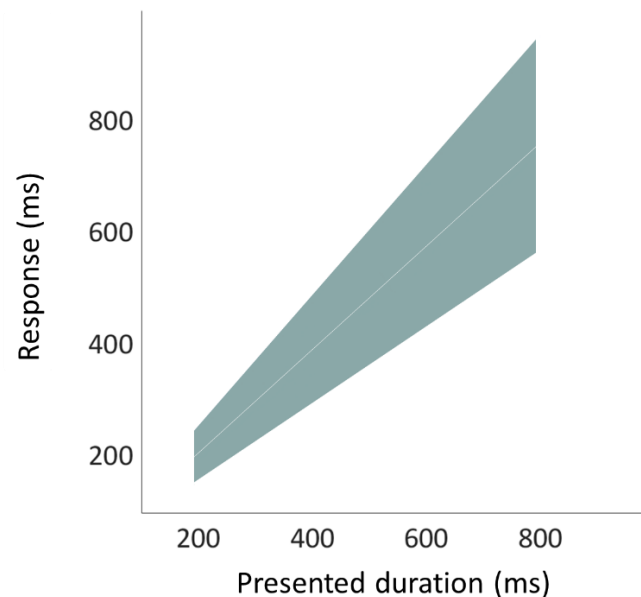


Figure 2.2. Scalar variability in time perception. *The variability of reported duration increases with increase in presented duration. The white line represents the mean of responses, and the shaded area the standard deviation of responses. As the presented duration increases, the variability of responses increases as well.*

Even though Weber's law provides a neat theoretical account of the variability of magnitude estimation, its relevance and applicability to the domain of time perception remain an open question. On the one hand, some studies have found that scalar variability is preserved for a range of durations with the coefficient of variation remaining constant from 0.5 to 1.3s (Wearden & McShane, 1988), and from 0.5 to 32s (Zeiler, Scott, & Hoyert, 1987, but only when data was aggregated across all participants – see Wearden, 1991). However, not all empirical findings support the predictions of Weber's law. Studies comparing the discrimination performance associated with a short compared to a long stimulus – 50 and 1000ms

(Rammsayer & Pichelmann, 2018), 200 and 2000ms (Lavoie & Grondin, 2004), 200 and 1000ms (Grondin, 2010) – have found that the Weber fraction was larger for the longer durations. Studies incorporating ranges of durations show similar results – the Weber fraction (and coefficient of variation) increases with duration in a range of supra-second intervals across duration discrimination and reproduction tasks (Grondin, 2012; see also Woodrow, 1930). However, studies incorporating even wider ranges of durations suggest that the positive linear relationship between absolute stimulus duration and discrimination does not hold for very short durations. When durations range from 50 to 3200ms (Getty, 1975), and 100 to 1000ms (Rammsayer, 2014) the relationship between the discrimination threshold and absolute duration follows a U-shaped function – as durations decrease below 200ms, the Weber fraction increases. Increase in Weber fractions for very short durations has also been reported in earlier research (Henry, 1948; Abel, 1972; Fetterman & Killeen, 1992). Overall, the results discussed here suggest that the original formulation of Weber’s law is applicable only for a limited range of durations. An alternative generalized formulation of Weber’s law has been proposed in which the variance of estimates is a combination of a variance component proportional to the stimulus magnitude and some constant residual variance (Getty, 1975). This version of the model seems to explain the increase of the Weber fraction for short durations well but still fails to provide a good fit for the increase in Weber fraction observed for long durations (see Killeen & Weiss, 1987 for another reformulation of Weber’s law). This suggests that time perception has different properties across different scales, even milliseconds to seconds, and raises the question of whether scalar variability should be considered a general property of interval timing at all.

Finally, assuming that duration estimation does follow Weber’s law at least partially, it is still widely debated what processes could give rise to these effects. On the one hand, following scalar timing theory it is possible that the perceptual system represents stimulus magnitude on a linear scale and estimates simply become noisier as stimulus magnitudes increase (Figure 2.3B; Gibbon 1977; Cantlon, Cordes, Libertus, & Brannon, 2009). On the other hand, it is possible that stimulus magnitude is represented on a logarithmic scale and scalar variability emerges when stimulus magnitudes are transformed into a linear scale for the sake of reporting (Figure 2.3A; Fechner, 1860; Petzschner & Glasauer, 2011; Ren et al., 2020). Which proposition provides a more accurate account of perception remains an open question.

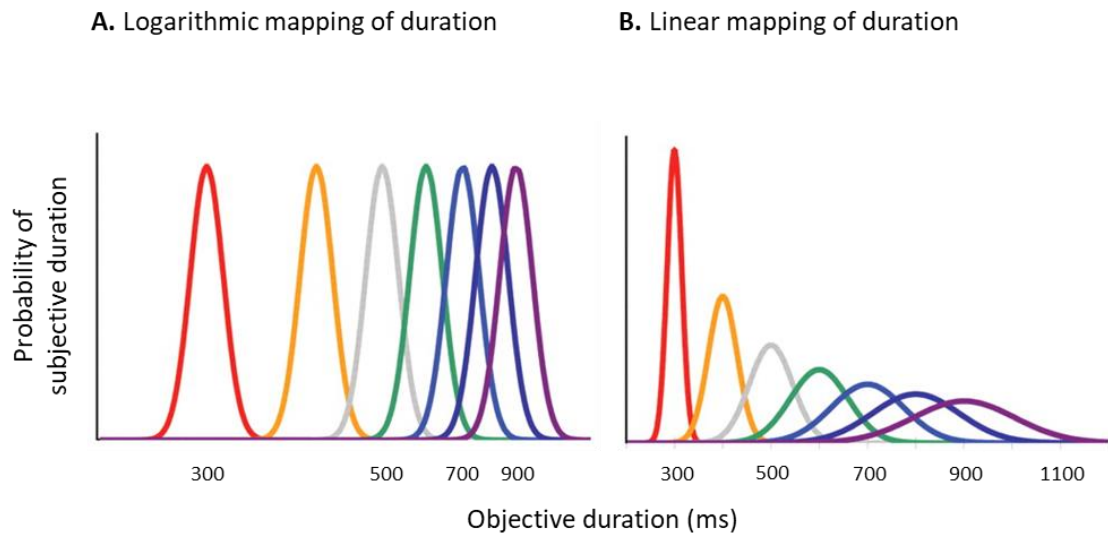


Figure 2.3. The two hypothesized representation of duration (or any magnitude characteristic). (A) If duration is encoded on a logarithmic scale, the variability of the representations is constant. As durations increase their representations become increasingly compressed which makes two long durations more difficult to distinguish than two short durations. (B) If duration is represented on a linear scale, durations are spaced at equal intervals, but the variability of their representations increases linearly with increase in the presented duration. Figure adapted from Cantlon et al. (2009).

2.3.2 Prior – building duration expectations

Our ability to define what kind of statistical information is reflected in the prior is limited because we do not have access to observers' subjective priors. Most studies analyze perceptual reports to indirectly determine the information the prior must contain for a given response to occur (more on this in the next section below). Going a step further, Acerbi, Wolpert, and Vijayakumar (2012) attempted to reconstruct the subjective prior distribution of each participant from their behavioural responses in a duration reproduction task. In most cases, the reconstructed priors were good approximations of the actual stimulus distributions – there was no evidence of differences between the mean, variance, and skewness of the participants' reconstructed priors and the presented stimulus distributions. However, the reconstructed prior differed from the real stimulus distribution when it came to kurtosis - participants' reconstructed priors typically had heavier tails than the stimulus distribution, and do not accurately reflect multimodality.

One key assumption of the Bayesian brain hypothesis is that priors are updated as new sensory information is encountered. Nonetheless, many studies using Bayesian perceptual models employ a fixed prior. When assuming a fixed prior, the shape of the prior remains the same from the first to the last trial in the experiment, and the prior does not change as a result of recent trial history. Some notable examples of studies using fixed priors include Jazayeri and

Shadlen (2010) and Acerbi et al. (2012). In contrast, in models assuming an updating prior, the prior changes on every trial as new information is encountered, similar to a Kalman filter. In updating priors, new and past trials can have equal weight in the computation of the prior (for example, Petzschner & Glasauer, 2011), or the weight of past trials can decay to mimic forgetting (for example, Ostwald et al., 2012). Studies find minor advantages of updating compared to fixed priors (Petzschner & Glasauer, 2011) and updating priors with forgetting compared to updating priors without forgetting (Ostwald et al., 2012). However, the reported differences in model performance are usually not substantial, which could be a by-product of specific task characteristics, specifically, the randomization of the stimuli. In a fully randomized trial sequence, either model will fit the data well because the structure of the immediate past trials will mimic the experiment's overall structure. In a random walk sequence, however, a Kalman model will fit the data better because the immediately past trials are not representative of the entire stimulus distribution (Glasauer & Shi, 2018). Ultimately, participants' subjective prior remains hidden, and researchers can only infer its most likely characteristics based on participants' reports.

2.3.3 Posterior – integrating measured and predicted duration

The main evidence that duration estimation, and magnitude estimation in general, arises from the integration of likelihood and prior comes from studies that investigate the effect of previously presented stimuli on perceptual reports. For example, it has been found that duration estimates exhibit serial dependence – judgements are biased towards the stimulus presented on the previous trial (Roseboom, 2019). On a more global scale, it has also been found that perceptual reports are biased not only towards the stimulus encountered in the immediate past but also towards the mean of the overall stimulus distribution. Durations longer than the mean are underestimated, and durations shorter than the mean are overestimated, while the mean represents a point of indifference – responses are equally under- and overestimated. This regression to the mean effect is range specific – the same physical value will be underestimated if it is larger than the mean value but overestimated if it is smaller than the mean value. Within the domain of duration estimation this central tendency bias is also known as Vierordt's law (Woodrow, 1934; Kowalski, 1943; Wearden, 2003; Lejeune & Wearden, 2009; Jazayeri & Shadlen, 2010, 2015; Cicchini, Arrighi, Cecchetti, Giusti, & Burr, 2012; Shi, Church, & Meck, 2013; Aagten-Murphy, Cappagli, & Burr, 2014; Murai & Yotsumoto, 2016; Shi & Burr, 2016; Roach, McGraw, Whitaker, & Heron, 2017; Barne et al., 2018; Rhodes, Seth, & Roseboom, 2018; Roseboom et al., 2019), and has been reported using

multiple different tasks. These include ready-set-go tasks, where durations are defined by two flashes or tones and participants have to produce a button press such that the interval between the second flash and the button press has the same duration as the interval between the first and the second flash (e.g., Jazayeri & Shadlen, 2010); interval reproduction, where participants have to produce an interval of the same duration as the stimulus by keeping the response button down (e.g., Acerbi et al., 2012); and direct estimation where participants report the perceived duration of the interval verbally (e.g., Experiment 3 in Penton-Voak et al., 1996), or using a visual analogue scale (e.g., Roseboom et al., 2019; Suárez-Pinilla et al., 2019). Furthermore, serial dependence and regression to the mean have been reported across other perceptual domains including serial dependence in orientation (Fischer & Whitney, 2014), numerosity (Cicchini et al., 2014), facial identity (Lieberman, Fisher, & Whitney, 2014), visual variance (Suárez-Pinilla et al., 2018), size (Hollingworth, 1910), and regression to the mean in loudness (Stevens & Greenbaum, 1966; Teghtsoonian & Teghtsoonian, 1978), movement planning (Hudson, Maloney, & Landy, 2007; Gallivan & Chapman, 2014), target-reach time (Chang & Jazayeri, 2018), linear and angular distance estimates (Petzschnner & Glasauer, 2011). The prevalence of such biases across different behavioural tasks and domains provides solid evidence that observers build and use predictions based on previously experienced sensory information to inform perception in the present moment.

Despite the wealth of research investigating regression to the mean effects in perception, the effect of the relative variance of the prior (or the likelihood) on perceptual reports remains a critical open question. Studies investigating how the variance of the prior affects estimates of location and force have found that when the prior has a smaller variance, responses become more optimal, and show greater regression towards the mean (Körding, Ku, & Wolpert 2004; Berniker, Voss, & Kording, 2010; Acerbi, Vijayakumar, & Wolpert, 2014). Focusing on the domain of time perception, Miyazaki et al. (2005) presented participants with sequences of three visual stimuli separated by equal inter-stimulus intervals and asked them to predict when the third stimulus would appear. In different blocks, the inter-stimulus intervals were drawn from a Gaussian distribution with a small or large variance. They found that participants' reports of stimulus arrival were more biased towards the mean presentation time when the distribution of inter-stimulus intervals had a small variance. However, in all these experiments the effect of variance was confounded by simultaneous manipulations of distribution range – distributions with large variance also had a larger range than distributions with small variance (Figure 2.4) – which prohibits strong conclusions about the specific

influence of prior variance on behavioural reports (under the mentioned Gaussian assumptions).

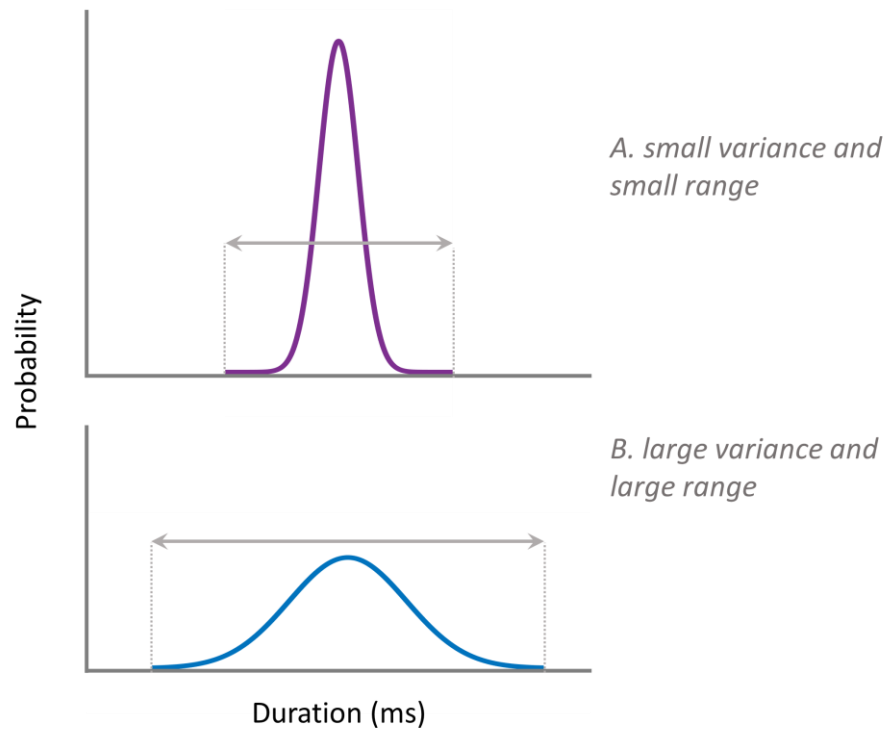


Figure 2.4. Confounding effects of range and variance. (A) Shows a “precise” distribution – it has a small variance and a small range, while (B) shows a “broad” distribution – it has a large variance and a large range. In experiments where variance and range are confounded, it is not clear the difference in which statistical feature is driving the differences in participants’ responses.

To date, very few studies have investigated how more complex distributional characteristics of the prior affect duration estimation, and the existing evidence shows that perceptual reports reflect the skewness but not multimodality of the prior distribution. For example, Ryan (2011) examined how duration reports differed if stimuli were drawn from a uniform distribution, a distribution skewed towards shorter durations and a distribution skewed towards longer durations. As in the classical regression to the mean effect, in the uniform distribution, short durations were overestimated, while long durations were underestimated. In contrast, when long distributions were overrepresented, participants’ responses to the long durations showed no bias, while all other durations in the range were overestimated. Similarly, when short durations were overrepresented, responses to short durations showed no bias, while all other durations were underestimated (also see Acerbi et al., 2012 for similar results). Nonetheless, all available evidence suggests that multimodality does not affect perceptual reports (Ryan 2011; Acerbi et al., 2012). Ryan (2011) found no differences in participants’ duration responses when the stimuli followed a uniform and a bimodal distribution, suggesting that observers do

not estimate bimodality. Similarly, Acerbi et al.'s (2012) found that when stimuli followed a bimodal distribution duration reports did not exhibit a strong bias towards the most probable stimuli, and participants did not assimilate the shapes of these distributions. The empirical results of both of these studies are consistent with the results from the prior reconstruction analysis reported by Acerbi et al. (2012) which suggested that participants' subjective prior did not represent bimodality – the statistical characteristics of stimulus distributions that are not reflected in the prior do not affect perception and perceptual reports.

Finally, evidence that perceptual reports reflect a process similar to Bayesian inference comes from studies investigating how sensory precision affects responses on duration estimation tasks. As discussed above, if the external properties of the stimulus are kept constant, the variance of the sensory likelihood will predominantly reflect participants' sensory precision which can be defined through Weber's law. Therefore, following the principles of Bayesian inference, greater sensory precision should result in a weaker bias towards the prior, but the research investigating this prediction is limited. The strongest available evidence that sensory precision affects the magnitude of perceptual biases comes from a study that found that participants' Weber fractions, estimated on the basis of a duration bisection task, were positively correlated with regression to the mean in their responses in a duration reproduction task (Cicchini et al., 2012). Importantly, these findings suggest that sensory precision is stable across tasks – sensory precision in duration bisection is related to sensory precision in duration reproduction. In addition, musicians tend to have smaller Weber fractions (i.e. greater sensory precision) than non-musicians (Grondin & Killeen, 2009), and studies have found that expert musicians tend to experience weaker regression to the mean than non-musicians across different duration perception tasks (Cicchini et al., 2012; Aagten-Murphy et al. 2014). These are promising findings providing evidence in favour of the prediction of Bayesian theories, but further research is needed to investigate the strength of the relationship between sensory precision and perceptual biases in the general population.

2.3.4 Conclusion

Positioning time perception within the framework of Bayesian inference can help us identify the gaps in our existing knowledge of what processes stand behind our ability to estimate durations. Despite recent advances in our understanding of the mechanisms involved in time estimation (Roseboom et al., 2019), we still don't know what determines the variability of temporal estimates or in what scale the brain encodes time. Furthermore, the research, investigating how the properties of the prior influence duration perception, has failed to

thoroughly investigate the effects of most statistical features apart from the mean. Most notably, no studies to date have examined the specific, non-confounded effects of prior variance on magnitude estimation in any modality, not just duration, which is surprising given that one of the key predictions of Bayesian integration related to the relative precision of the prior and the likelihood. Similarly, very few studies have investigated how sensory precision influences the integration of the prior and the likelihood, with the existing evidence focusing specifically on the difference between musicians and non-musicians.

2.4 Neurophysiology of time

In the previous section, I discussed the available behavioural evidence that time perception arises from a process akin to Bayesian inference. I will now focus on some of the neurophysiological correlates of time perception relevant to the integration of prior expectations and sensory inputs at the stage of decision making. Very briefly, the brain mechanisms behind the perception of time are still largely unknown. The literature surrounding this topic extends from pharmacological studies examining the roles of different neurotransmitters (Meck, 1996; Marinho et al., 2018), fMRI studies trying to determine how different brain areas are involved in duration estimation (Pouthas et al. 2005; Grondin, 2010; Sherman et al., 2020), and EEG studies investigating the time course of brain dynamics during timing (Macar & Vidal, 2004; Barne et al., 2018). Here I will focus on three physiological processes whose role in time perception and decision-making is still unclear – contingent negative variation (CNV), mismatch negativity (MMN), and dopamine release.

2.4.1 Contingent negative variation

One component of the evoked response potential that has been proposed to be associated with the perception of time is the contingent negative variation (CNV; Walter, Cooper, Aldridge, McCallum, & Winter, 1964). The CNV is observed as a response to a stimulus that signals the imminent presentation of another stimulus. It constitutes a prolonged component with a negative polarity that can reach several hundred milliseconds and can be observed after the early components of the ERP. Usually, the amplitude of the CNV ramps up following the presentation of the first stimulus, depending on the task it can reach a plateau, and finally goes back to baseline when the second stimulus is presented. The CNV has been observed with different stimulus pairs such as two sensory stimuli marking the start and end of an interval (Macar & Vitton, 1982), a sensory stimulus marking the start and a motor response marking

the end of the interval (Walker et al., 1964), and the onset and offset of a filled interval (Ng, Tobin, & Penney, 2011).

There is still a debate regarding what processes are reflected in the CNV (see Macar & Vidal, 2004). On the one hand, studies have found evidence that the CNV is related to the actual timing of the incoming sensory signal. For example, when participants were asked to reproduce an interval with a given duration, there was a positive correlation between participants responses and the amplitude of the CNV recorded while participants were producing their responses (Macar, Vidal, & Casini, 1999). The researchers also found that durations longer than the standard interval were associated with greater CNV amplitude than durations shorter than the standard interval in a discrimination task. As a result of such findings, proponents of the internal clock model have suggested that the CNV reflects the accumulation of pulses by the internal clock (Macar & Vitton, 1982; Macar et al., 1999; Casini & Vidal, 2011).

Nonetheless, many findings contradict the proposition that the CNV reflects only temporal accumulation. For example, in a duration bisection task, it was found that the amplitude of the CNV recorded in a given trial correlates with the duration presented on the previous trial – CNV amplitude was greater on trials preceded by longer durations (Wiener & Thompson, 2015). It was also reported that the CNV associated with the same physical duration differed depending on the context in which the stimulus was presented – the amplitude of the CNV was greater when the stimulus was the longest in the set compared to when the stimulus was the shortest in the set (Damsma, Schlichting, & van Rijn, 2020). In contrast to these findings, the accumulator in internal clock models is not affected by prior expectations. If the CNV reflected the accumulation process, then there should have been no effect of previously experienced stimuli on the CNV. These results instead suggest that the CNV reflects stimulus anticipation based on prior expectations. When the presented stimulus is shorter than expected, the ramping up of the CNV is cut short, and the amplitude remains low. Conversely, when the presented stimulus is longer than expected – the amplitude of the CNV increases until the anticipated event occurs. Finally, findings from temporal generalisation (Macar & Vidal, 2003) and bisection tasks (Ng et al., 2011) go one step further and suggest that the CNV could reflect the outcome of the decision-making process. In a bisection task it was found that for long stimuli, the amplitude of the CNV started to drop when the mean between the two anchors was reached rather than remain constant until the end of the stimulus (Ng et al., 2011). These results are evidence that the CNV reflects decision making because its time-course is

reminiscent of the strategy most observers would be expected to adopt in such a task – a stimulus will be identified as closer to the short anchor if it is shorter than the mean between the two anchors. Conversely, a stimulus will be identified as closer to the long anchor if it is longer than the mean. Overall, based on the existing evidence, it is more likely to assume that the CNV reflects stimulus anticipation and decision making, rather than merely stimulus processing, but more research is needed to provide conclusive results.

2.4.2 Mismatch negativity

Another component of the ERP that has been proposed to play a role in time perception is the mismatch negativity response (MMN), which is elicited by violations of regularity, and could therefore be related to the assimilation of statistical information in perceptual predictions. In the classical oddball paradigm used to measure MMN, participants are presented with a sequence consisting of the repetition of a stimulus known as the standard stimulus. Occasionally the stream of repeated stimuli is interrupted by the presentation of a deviant stimulus that differs from the standard. Originally, MMN was measured using EEG and was observed in the auditory domain as a negative component of the event-related potential (ERP) with fronto-central topography and latency between 100 to 200ms from the onset of an unexpected auditory tone presented in the absence of attention (Näätänen, Paavilainen, Rinne, & Alho, 2007). It is important to note that while both the CNV and MMN are affected by prior expectations, the information they encode is not the same (Macar & Vidal, 2004). In the domain of time perception, the CNV reflects the physical duration expected based on the prior and participants' decision regarding the duration of the stimulus. In contrast, the MMN reflects only how well that stimulus fits within the model of the world currently held by the brain. For example, consider two distributions with different ranges – one consisting of short durations (100-500ms), and one consisting of long durations (600-1000ms), but the same standard deviation (to provide a simple example we will assume that the brain encodes time in a linear scale). Observers would hopefully perceive the longest stimuli in the two distributions (500 and 1000ms) differently. As a result, the CNV elicited by the longest stimulus of the long distribution (1000ms) should have a greater amplitude than the CNV elicited by the longest stimulus of the short distribution (500ms). In contrast, the amplitude of the MMN elicited by these stimuli should be the same because, despite their physical differences, they are equally likely within the two distributions. The same principles hold if the brain encodes time on a logarithmic scale, only the ranges (when transformed to linear scale) would be different.

MMN has been observed across multiple sensory modalities including visual (Stephens, Kremláček, & Czigler, 2014), auditory (Näätänen et al., 2007), somatosensory (Strömmer, Tarkka, & Astikainen, 2014), and temporal perception (Chang, Seth, & Roseboom, 2017). In many experiments investigating temporal MMN, stimuli are presented after a fixed inter-trial interval which coupled with the repetition of the standard stimulus in oddball paradigms gives rise to a rhythmic structure (Tse & Penney, 2006; Roger, Hasbroucq, Rabat, Vidal, & Burle, 2009; Hsu et al., 2010). Consequently, the presentation of a deviant stimulus violates both expectations of duration and rhythm, and it is impossible to disentangle the neurophysiological responses associated with one from the other. More recently studies have reduced the confounding effect of rhythm by varying the duration of the inter-trial intervals, and have found that MMN can be elicited specifically as a result of violating expectations of duration (Chen, Huang, Luo, Peng, & Liu, 2010; Chang et al., 2017). Furthermore, studies have found that the amplitude of the MMN depends on variance and range of the stimulus distribution (Garrido, Sahani, & Dolan, 2013; Garrido, Teng, Taylor, Rowe, & Mattingley, 2016; Randeniya, Oestreich, & Garrido, 2019). For example, Garrido et al. (2013) found that the amplitude of the MMN in response to an unexpected stimulus was greater when the stimuli were drawn from the distribution with small variance and narrow range rather than a distribution with large variance and wide range. Such results provide compelling evidence that MMN could reflect the integration of prior expectations and sensory information as they are reminiscent of the effect of prior variance (and range) found in behavioural reports (e.g., Berniker et al., 2010; Miyazaki et al., 2005).

Despite the extensive research documenting the properties of the MMN, it is still unclear what neurological processes give rise to this component of the evoked response potential. The theories that have been most extensively examined relate MMN to neuronal sensory adaptation, change detection, model adjustment, and prediction error detection (Lieder, Daunizeau, Garrido, Friston, & Stephan 2013). From these theories, neuronal adaptation suggests that MMN reflects sensory processing, while the change detection, model adjustment and prediction error accounts relate MMN to detecting statistical irregularities with the last two proposing different roles for MMN within the Bayesian brain framework. Here I will focus on the neuronal adaptation, model adjustment, and prediction error accounts of MMN.

According to the neuronal adaptation hypothesis, MMN reflects early sensory processing rather than violations in regularity. Neuronal adaptation constitutes a decrease in a neuron's firing rate in response to repeated stimulation, and proponents of this theory suggest that

MMN is generated by differences in the fatigue of neuronal populations responding to a specific stimulus feature (May, Tiitinen, Ilmoniemi, Nyman, Taylor, & Näätänen, 1999; May & Tiitinen, 2010). The repeated presentation of a standard stimulus leads to a decrease in the response of the population of neurons that respond specifically to that stimulus, resulting in a decrease in the amplitude of the EEG signal. The neurons selective for the deviant stimulus are not adapted, and their activity leads to an enhanced response when the deviant is presented. While neuronal adaptation can provide a parsimonious account of MMN, sensory-specific adaptation cannot explain how MMN is elicited in paradigms that do not implement a repetition of a standard stimulus, or in paradigms in which the unexpected event is the repetition of a standard stimulus (Saarinen, Paavilainen, Schöger, Tervaniemi, & Näätänen, 1992; Garido, Kilner, Stephan, & Friston, 2009; Symonda et al., 2017). These findings suggest that even if neuronal adaptation has some role in the generation of MMN, MMN is unlikely to reflect just the processing of the sensory features of presented stimuli.

In contrast to the theory of neuronal adaptation of MMN, the model adjustment account and the prediction error account suggest that MMN reflects later stages of perceptual decisions rather than the initial sensory processing of the stimulus. According to both theories, the brain maintains a dynamic model of the environment that updates as new sensory information is encountered. The model adjustment hypothesis suggests that MMN reflects model updating in response to the incorporation of new information that does not match the predictions of the model (Winkler, Karmos, & Näätänen, 1996; Winkler & Czigler, 1998; Ostwald et al., 2013; Winkler, 2007). In contrast, the prediction error hypothesis suggests that MMN reflects the difference between the predicted sensory input and the incoming sensory information (Wacongne, Changeus, & Dehaene 2012; Stefanics et al., 2014). It is essential to note the distinctions between these two accounts of MMN. In the prediction error hypothesis, MMN reflects whether the predictive model accounts for the sensory input and therefore signals whether the model should be updated. This account differs from the model adjustment hypothesis where MMN reflects the difference between the prior and the posterior, or whether the predictive model *was* updated after the presentation of sensory information. Both theories can account for findings that neuronal adaptation cannot explain, and computational models of MMN based on the model adjustment hypothesis or the prediction error hypothesis outperform models based on neuronal adaptation (Ostwald et al., 2012; Lieder et al., 2013). Nonetheless, modelling studies to date have failed to provide convincing evidence in favour of one theory over the other (Lieder et al., 2013), and empirical studies have not yet found a way to distinguish these two theories.

To summarize, the evidence to date challenges the proposition that MMN reflects sensory processing, and supports the prediction error and model adjustment accounts which suggest that MMN reflects later stages of stimulus processing. However, empirical studies have not yet differentiated between these two theories, and it is still unclear whether MMN signals the updating of the predictive model or the discrepancy between predicted and actual sensory inputs.

2.4.3 Dopamine

The dopaminergic system has also been suggested to be involved in time perception (Coull, Cheng & Meck, 2011), but its exact role is still disputed. Proponents of scalar timing theory suggest that dopamine modulates the rate of the internal clock – the greater the level of dopamine in the brain, the higher the rate of the pacemaker (Matell & Meck, 2004; Meck, 1996). Early evidence in favour of this hypothesis suggested that in a temporal bisection task administering dopamine antagonists in rats (decrease dopamine levels) leads to an underestimation of time intervals, while administering dopamine agonists (increase dopamine levels) leads to an overestimation of time intervals (Maricq & Church, 1983; Meck 1983; Meck, 1986; Matell, King, & Meck, 2004). In addition, it was recently reported that the firing rate of neurons in the medial frontal cortex and the basal ganglia in non-human primates was flexibly adjusted based on the duration of produced temporal intervals (Wang, Narain, Hosseini, & Jazayeri, 2018), providing further supportive evidence for the involvement of dopaminergic activity in timing. Consequently, Wang et al. (2018) proposed a neural circuit model of action timing which was later expanded to account for the responses of human participants on different tasks requiring periodic production of actions (Egger, Lee, & Jazayeri, 2020).

Despite the research presented above, the existing pharmacological evidence in humans on the role of dopamine in time perception is mixed. Patients with conditions associated with low dopamine levels such as schizophrenia and depression have shown deficits in duration discrimination compared to healthy controls (Wahl & Sieg, 1980; Rammsayer, 1990). Patients with Parkinson's disease have also demonstrated deficits in duration estimation and production tasks (Pastor, Artieda, Jahanshahi, & Obeso, 1992; Lange, Tucha, Steup, Gsell, & Naumann, 1995). Crucially, administering l-dopa medication in Parkinson's patients increases dopamine and improves performance in duration estimation and reproduction (Pastor et al., 1992; Lange et al., 1995). These findings provide evidence that dopamine levels in the brain could be directly related to time perception – a decrease in dopamine levels leads to deficits in time perception tasks while an increase in dopamine levels resolves those deficits.

Nonetheless, other studies have found conflicting results which challenge the suggested role of dopamine. Non-medicated Parkinson's patients showed deficits on a bisection task when presented with supra-second but not sub-second intervals (Smith, Harper, Gittings, & Abernethy, 2007). Conversely, medicated patients performed no differently than controls on duration estimation of long intervals, 12-48s, but they still had deficits in duration discrimination of short intervals, 1-2s (Riesen & Schnider, 2001). It is also difficult to draw strong conclusions based on patient data alone as diseases have widespread, heterogeneous effects, which makes it difficult to relate a given result to one specific deficit in isolation. Unfortunately, manipulating dopamine levels in healthy controls has not provided conclusive results either. For example, administering a dopamine antagonist had an effect on duration discrimination, but administering a dopamine agonist or l-dope did not (Rammsayer, 1989a; Rammsayer, 1989b; Rammsayer, 1999).

Pharmacological manipulations also cannot be used to establish whether transient dopamine fluctuations alter duration perception on a trial-by-trial basis. Previous studies have found that dopamine release is positively correlated with the rate of spontaneous blinking in non-human primates (Karson, 1983; Groman et al., 2014). Nonetheless, many studies in healthy human participants have also failed to find a relationship between dopaminergic activity and spontaneous blinking (van der Post, de Waal, de Kam, Cohen, & Gerven, 2004; Mohr, Sandor, Landis, & Brugger, 2005; Dang et al., 2017; see Jongkees & Colzato, 2016 for a review), raising concerns about the reliability of this measures. Despite the controversy in the literature, Terhune, Sullivan, and Simola (2016) decided to investigate how trial-by-trial fluctuations in dopamine levels alter duration perception using spontaneous blinks as a proxy for dopamine activity. They found that in a duration bisection task, observers overestimated durations on trials preceded by blinks. The effect was observed across intervals presented in both the visual and auditory domain. Nonetheless, it was recently found that pre-trial spontaneous blinking did not affect duration estimation of naturalistic videos (Suárez-Pinilla et al., 2019), and raises further questions regarding the potential involvement of dopamine in time perception, and the relationship between spontaneous blinking and duration estimation. Recently, Mikhael and Gershman (2019) suggested that the conflicting evidence presented above can be reconciled by reshaping the role of dopamine in timing through its role in signalling reward-prediction errors – an unexpected reward leads to an increase in dopamine levels (and increase in the rate of the pacemaker) while an omission of an expected reward leads to a decrease in dopamine levels (and decrease in the rate of the pacemaker).

An alternative possibility for the role of dopamine in duration estimation tasks is that dopamine does not reflect a duration-related component of the decision process, but is related to more general processes like response biases, action selection, and motivation which all influence performance on perceptual tasks. One study found that stimulation to the basal ganglia, which results in increased release of dopamine, led to an increase in the probability of affirmative responses in a change detection task (both hits and false-alarms, Wang, Rangarajan, Gerfen, & Krauzlis, 2018). Such results suggest that changes in dopamine concentration could act by shifting the decision criterion – evidence exceeding the decision criterion leads to an affirmative response (a change occurred), while evidence that falls below the criterion leads to a negative response (a change did not occur). Lowering the decision criterion would result in more changes being successfully detected, but also in more trials lacking a change being falsely labelled to contain a change. Criterion shifts can also potentially explain Terhune et al.'s (2016) results – an increase in dopamine may lead to a transient increase in propensity to respond “long” without changing the available sensory evidence. This could also explain why no effect of pre-trial spontaneous blinking has been observed in a duration estimation task, which is potentially less subject to such biases (see Suárez-Pinilla et al., 2019). In addition to biasing responses, dopamine concentrations have also been found to reflect action choice (Howard, Li, Geddes, & Jin, 2017). Rats were presented with two levers, one to the left and one to the right, and trained to press the left lever in the first half of the trial and the right lever in the second half of the trial. The researchers found that rats showed high dopamine concentration in the first half of the trial, and low dopamine concentrations in the second half of the trial, suggesting that dopamine levels were specifically reflecting the action chosen by the animal. Finally, it has been proposed that shifts in duration perception due to changes in dopamine concentration could be related to changes in motivation (Wise, 2004; Balci, 2014). For example, Soares, Atallah, & Paton (2016) found that suppressing the activity of dopaminergic neurons in rats resulted in flattening of the psychometric curve in a duration discrimination task. Based on such results, it is difficult to view Terhune et al.'s (2016) findings only through the interpretation that dopamine speeds up the (hypothetical) internal clock.

It is also important to consider that the effect of spontaneous blinking on duration estimation may not reflect changes in dopaminergic activity. Blinking leads to gaps in perceptual content with variable duration (VanderWerf, Brassinga, Reits, Aramideh, & Ongerboer de Visser, 2003). Spontaneous blinking also leads to decreased sensitivity to retinal stimulation (Volkman, Riggs, Moore, 1980) and is associated with decreased activation in the visual cortex (Bristow,

Frith, & Rees, 2005; Golan et al., 2016). It was also found that in a duration bisection task, blinks on the current trial led to an underestimation of durations for visual but not auditory stimuli (Grossman et al., 2019). These results can be explained by the theory that duration estimation is based on perceptual content (Roseboom et al., 2019) – blinking reduced perceptual content in the visual but not auditory domain, and in turn, leads to an underestimation of visually presented durations. Within this framework, it is also possible to explain Terhune et al.'s (2016) results without referring to dopamine at all. Blinking leads to duration underestimation and simultaneously increases the uncertainty associated with the temporal estimate made at the time of blinking. As a result, intervals experienced immediately after a blink might be biased away from the estimate made on the previous trial, leading to an overestimation of the presented duration.

2.4.4 Conclusion

Different theoretical accounts and empirical studies have related CNV, MMN, and dopamine concentration to sensory or decision-making aspects of perception. On the one hand, CNV has been proposed to reflect the duration of the presented stimulus, MMN – to arise as a result of low-level sensory adaptation, and dopamine release – to regulate the speed of the internal clock. On the other hand, many studies provide evidence that these processes are involved in later stages of processing with CNV reflecting the outcome of the decision process, MMN – the prediction error between priors and sensory inputs or the integration of sensory inputs into the brain's generative model, and dopamine release – shifts in response bias. Nonetheless, the debate in the literature is far from resolved, and more research is needed to establish the role of these processes in time perception and Bayesian inference.

2.5 Bayesian inference outside of awareness

All studies discussed so far have investigated the mechanisms of perceptual inference under conditions of full awareness. Nonetheless, multiple lines of research provide evidence that some sensory processes can still occur when observers are not aware of the encoded sensory information. For example, adaptation to motion (OShea & Crassini, 1981; Roumani & Moutoussis, 2020), and orientation (Wade & Wenderoth, 1978; He & Macleod, 2001) still occurs even if the stimuli are suppressed from awareness. In contrast, adaptation to more complicated features such as face expression (Yang, Hong, & Blake, 2010) and identity (Moradi, Koch, & Shimojo, 2005) require that the adapting stimulus be visible.

Nonetheless, not many studies have investigated if subjective priors and the integration of prior expectations with sensory input require awareness. Suárez-Pinilla (2018) used continuous flash suppression (CFS) to examine if serial dependence in orientation perseveres outside of awareness. CFS is a masking paradigm in which a high-contrast flashing stimulus, the mask, presented to the dominant eye suppresses a low-contrast target presented to the non-dominant eye from entering awareness (Tsuchiya & Koch, 2005). In Suárez-Pinilla's (2018) study a grating of variable orientation was presented to the participants' non-dominant eye, while on half of the trials a dynamic high-contrast mask was presented to the dominant eye to suppress the target stimulus from awareness. It was found that visible stimuli affected responses on subsequently presented gratings, while suppressed stimuli did not, suggesting that serial dependence does not occur outside of awareness (also see Kim, Burr, Cicchini & Alais, 2020 for further support of these findings). In contrast to these results, Suárez-Pinilla (2018) found that representations of central tendency could take into account masked stimuli. The orientations of the presented stimuli varied between -60 to +60 degrees. Across two conditions, either leftward (-60 to -24 degrees) or rightward orientations (24 to 60 degrees) were masked, with the rest of the orientations being visible, with orientations between -12 and +12 degrees visible in both conditions. If regression to the mean depended only on the orientation of the visible stimuli, responses to the overlapping orientations would be biased towards leftwards orientations when rightward orientations were masked and vice versa. In contrast to this prediction, the researchers found strong evidence that there was no difference in participants' responses to the overlapping stimuli between the two conditions, suggesting that the masked stimuli were taken into account when building a global representation of central tendency. Nonetheless, the authors list several methodological concerns which bar very strong conclusions, most notably that the short duration of the experimental blocks may not have been enough for participants to form accurate representations of the prior distribution.

Studies have also investigated whether the brain tracks the statistical regularities of sensory input using EEG to measure the MMN elicited by unexpected stimuli at different levels of awareness. Studies have reported that MMN can be measured during sleep, suggesting that awareness is not necessary for the brain to track statistical regularities. In one study, Atienza, Cantero, & Gómez (1997) found that a classical auditory oddball sequence elicited MMN during both wakefulness and REM sleep (see Loewy, Campbell, & Bastien, 1996 for supporting results). Extending these findings, Atienza and Cantero (2001) reported that after behavioural training, changes in complex sound sequences elicited an increase in MMN amplitude during

both wakefulness and REM sleep. Furthermore, these studies did not find any significant differences in the characteristics of the MMN between wakefulness and REM sleep, suggesting that the same processes of regularity tracking operate across different levels of consciousness. Other studies have reported that MMN could not be elicited by changes in auditory pitch during non-REM sleep (Winter, Kok, Kennerly, & Elton, 1995; Loewy et al., 1996), suggesting that the perceptual system does not track the statistical characteristics of stimuli in the same way across the entire spectrum of consciousness. In another line of research, studies have investigated whether MMN persists outside of awareness by manipulating the visibility of stimuli. Jack et al. (2017) used binocular rivalry to make the deviant stimulus in a visual MMN oddball paradigm invisible. Participants were presented with gratings of different orientation to the left and the right eye, and standard and deviant stimuli differed in contrast – the deviant had a lower contrast than the standard. The deviant was presented either to both eyes, only to the dominant eye, or only to the non-dominant eye. The researchers found that unperceived deviants presented to the non-dominant eye still elicited a change in MMN amplitude compared to the standard stimuli. However, an important limitation of this study, which the authors acknowledge, is that the changes in MMN amplitude could be entirely explained by the contrast differences between the standard and deviant stimuli, as stimuli with lower contrast generally elicit evoked responses with lower amplitudes than stimuli with high contrast. Another study using stimuli not confounded by changes in contrast showed that visual MMN could not be evoked in the absence of perceptual awareness (Flynn, Liasis, Gardner, & Towell, 2016). Flynn et al. (2016) used visual stimuli that differed in orientation and used forward and backward masking to make both the standard and deviant stimuli invisible. Using this design, Flynn et al. (2016) found that when participants could not detect the presentation of a stimulus, there were no differences in the ERPs elicited by standard and deviant stimuli, suggesting that the perceptual system cannot track the statistical regularities of visual stimuli that do not enter perceptual awareness. Overall, the contradictory results cited by studies investigating if the brain tracks statistical regularities in the absence of awareness necessitate further research into this topic.

2.6 Gaps in the existing literature

Extensive research has investigated whether perception follows the principles of Bayesian inference. Studies have mainly examined how the mean and range of the prior distribution influence perceptual reports, but the effects of distributional variance on perception and brain activity are still largely unknown. To date, very few studies have examined how the variance of stimulus distributions alters the influence of prior expectations on stimulus processing and

behaviours reports, and the effect of variance has always been confounded by simultaneous manipulations of distributional range. The hypothesis that reports should be biased towards the mean of the stimulus distribution more when the distribution has a small variance is an important assumption of the Bayesian brain hypothesis, which needs to be thoroughly examined to ascertain whether human perception is truly Bayesian. The neurophysiological processes that could allow the brain to form representations of the statistical properties of the prior distribution and to form decisions based on these summary representations are also still only vaguely understood. Bayesian models also predict that the effect of the prior on perceptual reports depends on the variance of the sensory likelihood. Sensory precision is one of the key variables which determine the variance of the sensory likelihood, but to date, the research examining whether perceptual biases are influenced by the sensory precision of the observer has been limited. Finally, we still do not know to what extent awareness is necessary to form and update perceptual predictions. Some studies suggest that observers can form prior expectations in the absence of awareness, but it is still not clear how sophisticated these predictions can be under limited resources. For example, it is possible that without awareness the perceptual system can form only limited priors that do not take into account complex statistical information or are otherwise constrained to basic perceptual dimensions such as auditory pitch rather than abstract dimensions such as duration. Situating these questions within the domain of time perception adds a layer of complexity and challenges because although the way people track time has been the subject of scientific research for many decades, there are still many basic characteristics of time perception that are not as well understood as the architecture of more primary domains such as the processing of visual and auditory features of objects.

Chapter 3. Bayesian models

In Chapters 4 and 5, we use different versions of an ideal observer model to estimate the required sample size for each experiment and to form numerical predictions regarding participants' responses in the duration perception tasks. All versions of the model produce predicted behavioural responses based on participants' internal estimate of the stimulus distribution (i.e., the subjective prior), the sensory likelihood associated with the stimulus presented on every trial, and the subjective error mapping between the presented duration and the responses. Model predictions of duration reproduction (Chapter 4 and Chapter 5, Experiment 1) are also based on the motor likelihood associated with the selected motor command, while predictions of direct duration reports (Chapter 5, Experiment 2) are affected by the report noise associated with giving responses on a visual analogue scale. Motor noise and response noise are implemented in the same way across the models, but we use different terms to reflect the difference in the processes involved with the two modes of response generation.

3.1 Model 1: Fixed prior, durations represented on a linear scale

In Chapter 4, we modelled participants' responses in a duration reproduction task using the model detailed by Acerbi et al. (2012). The below description of this model is from Baykova et al. (2019). The components of the model are illustrated in Figure 3.1.

On every trial, participants are presented with a stimulus x drawn from the probability distribution $p(x)$. We assume that participants would form an internal representation y of the presented stimulus and that this internal representation is corrupted by sensory noise w_s . Therefore, the sensory likelihood in the model defines the probability distribution of y , and is approximated by a Gaussian probability distribution centred at the duration of x with a standard deviation equal to w_s , $p_s(y|x; w_s)$. Next, participants produce a response r by generating a motor command with duration u , which is corrupted by motor noise w_m . The motor likelihood in the model defines the probability distribution of the response r , and is approximated by a Gaussian probability distribution centred at the duration of the motor command u with a standard deviation w_m , $p_m(r|u; w_m)$.

After participants produce a response, they receive signed feedback which specifies the error between the response r and the stimulus x , which is computed as $f(x,r)=r-x$. The model assumes that participants build a subjectively expected loss function that is the quadratic form,

$\widetilde{f}^2(x, r)$, of the objective loss function $f(x, r)$. Finally, throughout the experiment, participants are expected to build an internal representation of the stimulus probability distribution. This subjective prior $q(x)$ is represented as a Gaussian approximation of the objective stimulus distribution in each condition.

The first part of the model, the internal model (Equations 9 to 11), determines the optimal action, $u^*(y)$, which is the motor command u that minimizes the subjectively expected loss between the stimulus x and the response r :

$$u^*(y) \propto \arg \min_u \int p_s(y|x; w_s) q(x) p_m(r|u; w_m) \widetilde{f}^2(x, r) dx dr \quad (9)$$

Where the subjective loss function is defined as:

$$\widetilde{f}^2(x, r) = (r - x)^2. \quad (10)$$

Equation (9) then becomes:

$$u^*(y) \propto \arg \min_u \sum_{i=1}^{1162} p_s(y|x_i; w_s) q(x_i) [w_m^2 + (u - x_i)^2]. \quad (11)$$

Crucially, because the participants do not know which stimulus is presented, the optimal action must minimize the subjective loss across all possible stimulus durations.

The second part of the model, the generative model – Equation (12), generates the probability distribution of responses $p(r)$ based on the sensory likelihood of the presented stimulus x and the motor likelihood of the optimal response $u^*(y)$:

$$p(r|x; w_s, w_m) \sim \int p_s(y|x; w_s) p_m(r|u^*(y); w_m) dy. \quad (12)$$

The probability distribution of responses $p(r)$, the posterior, has a mean and variance equal to:

$$\mu_{Posterior} = \frac{x \times w_s^2 + u^*(y) \times w_m^2}{w_s^2 + w_m^2}, \text{ and } \sigma_{Posterior}^2 = \frac{w_s^2 \times w_m^2}{w_s^2 + w_m^2} \quad (13)$$

Finally, the model-predicted response is derived by integrating the posterior probability distribution of responses over the possible values:

$$r \sim \int p(r|x; w_s, w_m) dx. \quad (14)$$

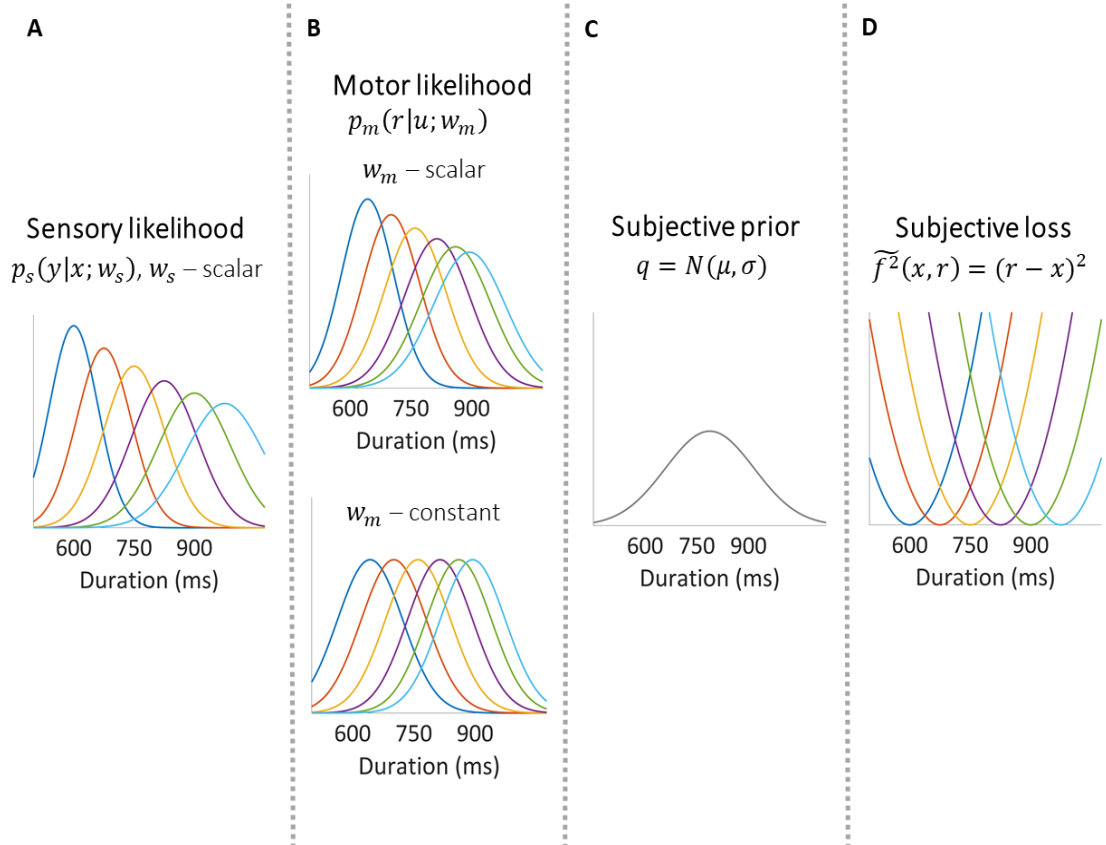


Figure 3.1. Components of Model 1. **(A)** The sensory likelihood is a Gaussian distribution centred at the presented stimulus with variance which is expected to be proportional to the duration of that stimulus. **(B)** The motor likelihood is a Gaussian distribution centred at the duration of the selected motor command with variance that is expected to be either proportional to the duration of the motor command (top), or constant (bottom). **(C)** The subjective prior is a Gaussian distribution with mean and variance equal to the mean and variance of the objective stimulus distribution. **(D)** The subjective loss function is represented as the squared loss function. This figure is from Baykova et al. (2019).

3.2 Model 2: Fixed prior, durations are represented on a logarithmic scale

In Chapter 5, Experiment 1, we used a modified version of the model above to predict participants' responses in a task where they have to reproduce the duration of visually presented stimuli. Following the evidence in the literature that estimates of stimulus magnitude are better represented on a logarithmic scale (Petzschner & Glasauer, 2011; Petzschner, Glasauer, & Stephan, 2015), in the modified version of the above model, hidden variables and computations are defined on a logarithmic scale. As in the original model, on every trial, participants are presented with a duration x which follows the probability distribution $p(x)$ in linear space. The internal representation of the stimulus, y_l , is transformed to a logarithmic scale, and the sensory likelihood is defined as a log-normal probability distribution, $p_s(y_l|x_l; w_{sl})$:

$$x_l = \ln \frac{x}{\sqrt{1 + \frac{w_s^2}{x^2}}}, \text{ and } w_{sl} = \sqrt{\ln(1 + \frac{w_s^2}{x^2})} \quad (15)$$

where w_s is the sensory noise parameter defined in linear space.

The motor likelihood is also defined as a log-normal distribution, $p_m(r_l|u_l; w_{ml})$ – the observable variable r_l and the hidden variable u_l are represented on a logarithmic scale, and the standard deviation is defined as:

$$w_{ml} = \sqrt{\ln(1 + \frac{w_m^2}{\exp(u_l)^2})} \quad (16)$$

where w_m is the motor noise parameter defined in linear space.

After providing a response, participants receive feedback which is computed as the difference between the duration of their response and the duration of the stimulus in linear space: as $f(x, r) = (r - x)$. The subjectively expected loss function is, however, computed on a logarithmic scale as the squared difference between the estimated response transformed into log-space and the stimulus transformed into log-space: $\tilde{f}^2(x_l, r_l) = (r_l - x_l)^2$. Finally, the subjective prior $q(x_l)$ is a Gaussian approximation of $p(x)$ transformed into log-space. The objective stimulus distribution $p(x)$ has a mean of a , and a standard deviation of b . From there, $q(x_l)$ is a log-normal distribution centred at a_l , with a standard deviation b_l :

$$a_l = \ln \frac{a}{\sqrt{1 + \frac{b^2}{a^2}}}, \text{ and } b_l = \sqrt{\ln(1 + \frac{b^2}{a^2})}. \quad (17)$$

The computations involved in the two stages of the model described above, the internal model and the generative model, therefore take place in log space:

$$\text{Internal model: } u_l^*(y_l) \propto \arg \min_{u_l} \sum_{i=1}^5 p_s(y_l|x_{li}; w_{sl}) q(x_{li}) [w_{ml}^2 + (u_l - x_{li})^2]; \quad (18)$$

$$\text{Generative model: } p(r_l|x_l; w_s, w_m) \sim \int p_s(y_l|x_l; w_{sl}) p_m(r_l|u_l^*(y_l); w_{ml}) dy_l. \quad (19)$$

The posterior estimate (the integral of the posterior distribution) is transformed back into linear scale to provide the observable response:

$$r \sim \exp(\int p(r_l|x_l; w_{sl}, w_{ml}) dx_l). \quad (20)$$

3.3 Model 3: Updating prior, durations are represented on a logarithmic scale

In Chapter 5, Experiment 2, we use a Bayesian ideal observer model with an updating prior to predict regression to the mean in a duration estimation (rather than reproduction) task. As in Model 2, hidden variables are represented on a logarithmic scale and responses are translated to linear scale for reporting. The only difference between the two models is that in Model 3, participants' subjective prior updates after every trial by a Kalman filter.

On each trial participants are presented with a duration that is transformed to log-space, x_l . The mean of the sensory likelihood is equal to the presented duration x_l corrupted by some random noise, n , which is randomly drawn from the normal distribution $N(0, w_{sl})$. The standard deviation of the sensory likelihood is equal to w_{sl} . The subjective prior $q(x_l)$, changes on every trial. On the first trial, we assume that the mean and standard deviation of the prior, a_l and b_l , are equal to the presented duration and the sensory noise associated with that duration, x_l and w_{sl} respectively. On subsequent trials, the mean and standard deviation of the prior update depending on the Kalman gain, which is calculated as:

$$K_i = \frac{b_{li}^2}{b_{li}^2 + w_{si}}, \quad (21)$$

where i designates the current trial index. We also calculate the prediction error between the mean of the sensory likelihood, $x_l + n$, and the mean of the prior, a_l , for trial i as:

$$PE_i = (x_{li} + n_i) - a_{li} \quad (22)$$

Then, the mean and standard deviation of the prior for the next trial update to become:

$$a_{li+1} = a_{li} + K_i \times PE_i, \text{ and } b_{li+1} = \sqrt{(1 - K_i) \times b_{li}^2}. \quad (23)$$

As in Model 2, the motor likelihood is defined as a log-normal distribution, $p_m(r_l|u_l; w_{ml})$. The only difference between the definition of the motor likelihood between Model 2 and Model 3 is that here we refer to the standard deviation of the motor likelihood, w_{ml} , as report noise rather than motor noise. In duration reproduction (Models 1 and 2), participants' ability to provide a given response depends on the precision of the produced motor command. In contrast, direct estimations of duration where participants report the presented duration on a visual analogue scale (Model 3), would not depend as much on the precision of participants' motor command but mostly on participants' constraint of what deviation from the internal representation of the stimulus is acceptable.

The feedback and the associated subjectively expected loss function, the internal model, the generative model, and the selection of the final response are the same as in Model 2.

3.4 Model parameters free to vary

Across all models, sensory noise (w_s) and motor/response noise (w_m) are free to vary. In Chapter 4, the sensory and motor noise parameters associated with each duration were determined separately for each participant based on their performance on two additional behavioural tasks – a 2-interval forced-choice task and a reaction time task. In Chapter 5, sensory noise was also determined empirically, while the value of the motor/report noise parameter for each participant was fitted using the Metropolis-Hastings algorithm (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953; Hastings, 1970).

Chapter 4. Influence of prior distribution variance on time perception

Bayesian models of perception propose that the extent to which perception is affected by prior expectations depends on the precision or variance of the prior distribution – more precise priors will bias perception more strongly than broader priors. However, evidence for this in human behaviour is limited. It has been shown that reproductions of duration are more biased towards the prior distribution of presented stimuli when that distribution is peaked rather than uniform. However, in this case, the two distributions also differed in central tendency and skewness, making it difficult to disentangle the specific effect of distributional variance on perception. We conducted two duration-reproduction experiments in which participants were presented with durations drawn from a precise or broad prior. Experiment 1 (N=7) was a replication of an experiment by Acerbi et al. (2012), and participants were presented with a skewed or uniform distribution with the same range. In Experiment 2 (N=8), the durations were drawn from a uniform or Gaussian distribution that had the same mean, range, and skewness. By keeping all other statistical properties constant across stimulus distributions, we could directly investigate the influence of distributional variance on time perception. In Experiment 2, we found that participants' responses were more biased towards the prior in the Gaussian compared to the uniform condition, supporting the prediction of a Bayesian ideal observer model. However, in Experiment 1, we found a dissociation between the model's predictions and participants' responses in the skewed condition as the data showed evidence of both attractive and repulsive biases. To conclude, our results provide direct evidence in favour of one of the most fundamental assumptions of the Bayesian brain framework but also show the explanatory limits of Bayesian models thus calling for a greater focus towards developing a general framework that resolves the apparent contradiction between opposing perceptual biases.

4.1 Introduction

Subjective sensory experience is influenced by our prior knowledge about the world (e.g., Ma & Jazayeri, 2014). One efficient method of summarizing the regularities of the world would be to track the statistical characteristics of sensory information (Khayat & Hochstein, 2018) over time. These characteristics might include, for example, what stimulus appears most often, and how variable or reliable the sensory inputs are. In recent years, it has become popular to describe this influence of previous experience on subsequent perception using 'Bayesian'

frameworks wherein perception is viewed as a probabilistic inference reflecting the integration of sensory inputs with prior expectations about their potential causes (Friston, 2010; Clark, 2013; Hohwy, 2013). Typical computational models of this inference process represent sensory inputs (likelihoods) and prior expectations as probability distributions (Knill & Pouget, 2004; Pouget et al., 2013; Ma & Jazayeri, 2014). However, it is not clear how exactly prior expectations are represented in the brain (Walker, Cotton, Ma, & Tolia, 2018; Beck et al., 2008). In this study, we will investigate the assumption of common Bayesian accounts of perception that the variability of prior experience strongly determines its subsequent influence on perception. To do so, we will examine how the variability of the distribution of recently experienced temporal durations influences human participants' subsequent reproductions of duration.

To effectively utilise information from prior sensory experience, human observers must be able to track the variability of particular sensory inputs over time. Perceptual reports are well-known to reflect some statistical characteristics of the distribution of previous sensory experience. It has long been established that perceptual reports across different sensory domains and tasks are biased towards the mean of previously presented stimuli (Lejeune & Wearden, 2009; Hollingworth, 1910; Petzschner et al., 2015). Values greater than the mean are underestimated, and values smaller than the mean are overestimated. This 'regression to the mean' effect is range specific – the same physical stimulus value will be underestimated if it is greater than the mean but overestimated if it is smaller than the mean – and such biases have been reported in a range of domains, including time perception (e.g., Jazayeri & Shadlen, 2010; Cicchini et al., 2012; Roseboom, 2019; for a full review see *Literature review*, p.30).

Despite the wealth of research investigating regression to the mean effects in perception, the extent to which observers internalize information about the variance of stimulus distributions when forming predictions remains a critical open question. A common interpretation of Bayesian brain frameworks proposes that prior expectations and sensory likelihoods are represented as probability distributions, often specifically assuming they are well-approximated by a Gaussian distribution (for example, Friston & Kiebel, 2009; Pouget et al., 2013; Darlington, Beck, & Lisberger, 2018). Under this assumption, the influence of prior expectations on perception depends on the variance of the prior– keeping sensory information constant, the smaller the variance of the prior, the more perception is biased towards the prediction (Figure 4.1). Evidence broadly consistent with this prediction has been provided by studies investigating how the variance of prior distributions affects estimates of the location of hidden targets (Berniker et al., 2010; Acerbi et al., 2014), performance on a force-matching

task (Körding, et al., 2004) and predictions of stimulus arrival time (Miyazaki et al., 2005). For example, Berniker et al. (2010) and Körding et al. (2004) found greater regression to the mean in participants' responses when the prior distribution had a small variance, while Acerbi et al. (2014) found that responses were more optimal when the prior distribution had a smaller variance. It has also been suggested that features of brain activity, specifically the mismatch negativity evoked response, reflect differences in distributional variance and range (Garrido et al., 2013; Garrido et al., 2016; Randeniya et al., 2019). However, in all these experiments the effect of variance was confounded by simultaneous manipulations of distribution range – distributions with large variance also had a larger range than distributions with a small variance – which prohibits strong conclusions about the specific influence of prior variance on behavioural reports (under the mentioned Gaussian assumptions).

Focusing specifically on duration estimation, evidence that human observers may assimilate the variance of stimulus distributions when reporting the length of temporal intervals was provided by Acerbi et al. (2012). Manipulating the variance of a stimulus distribution entails manipulating the probability with which different stimulus levels are presented during the experiment. One of the experiments reported by Acerbi et al. (2012), Experiment 2, revealed that manipulating the probability of presentation of the different durations influenced time judgements. When durations were drawn from a uniform distribution, responses were biased towards the mean of the stimuli. When durations were drawn from a distribution in which a short stimulus was presented more frequently than any other stimulus duration (resulting in a short mean duration), participants underestimated the duration of all stimuli (responses were biased towards the stimulus with high probability). These results suggest that behavioural responses were affected by the probability of different stimulus levels and by extension, their variance. However, by increasing the probability of a stimulus that did not coincide with the mean of the uniform distribution, Acerbi et al. (2012) also changed the mean and skewness of the distribution. In Acerbi et al.'s (2012) study, the simultaneous manipulation of several distributional properties across conditions does not pose a challenge for the validity of their inference, because their main objective was to examine, computationally, whether

participants' reconstructed priors resembled the statistical properties of the objective stimulus distributions in different conditions.

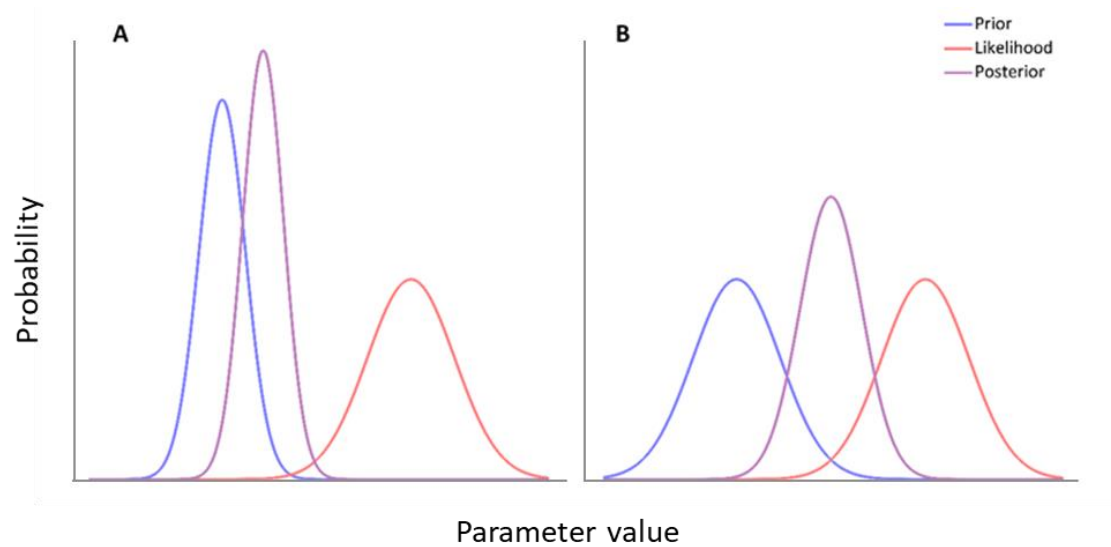


Figure 4.1. Predicted effect of prior variance on perception. According to Bayes rule, the posterior (purple) results from the integration of the prior (blue) and likelihood (red). When the prior has a small variance (**A**), the resulting posterior is biased towards the prior more than when the prior has a large variance (**B**).

However, investigating the specific effect of distributional variance on behavioural performance necessitates manipulating only the variance of the presented distributions while keeping other statistical properties the same across conditions. In this case, the concurrent differences in mean, variance, and skewness between distributions make it impossible to determine what proportion of the behavioural difference between conditions could be attributed to the manipulation of variance alone.

Here we report a study extending the findings of Acerbi et al. (2012) by specifically testing the influence of the variance of stimulus distributions on perceptual reports, avoiding confounding changes in other distributional characteristics. First, we replicated the key experiment from Acerbi et al.'s (2012) study examining the role of changing the mean and variance of the presented stimulus distribution. This previous experiment provides the strongest available evidence that human observers assimilate information about the variance of underlying stimulus distributions in the context of temporal perception, and it is, therefore, important to validate their findings. In a second experiment, we will directly investigate the effects of variance on behavioural responses. As in Acerbi et al.'s (2012) original study, participants were asked to reproduce the duration of presented temporal intervals manually. In one condition, the intervals were drawn from a truncated uniform distribution, and in a second condition from a truncated Gaussian distribution. Importantly, in both conditions the distribution had

the same mean, range, and skewness, differing only in distribution variance. We hypothesised that duration judgements would be more strongly biased towards the mean of the stimulus distribution when the stimuli follow a Gaussian compared to a uniform distribution. This prediction follows from the assumption that sensory priors and likelihoods are well-approximated by Gaussian distributions.

The experiments detailed here provide a direct way to test whether human perception, in the domain of time judgements, behaves in a way consistent with Bayesian processing. Furthermore, the present research can more generally further our understanding of how human observers extract summary statistics from sensory stimuli presented over time and help us better understand how prior expectations are used in perceptual processing.

4.2 Method

4.2.1 Pre-registration and data availability

The stage 1 registered report accepted by *Royal Society Open Science* is available at the Open Science Framework (<https://osf.io/75p42/>). The data is available here: <https://osf.io/75p42/files/> (this includes data from the simulations to determine sample size, and the behavioural data).

4.2.2 Design

Across two proposed experiments, we investigated how the shape of the distribution from which durations are drawn influences the mean and standard deviation of human participants' responses. In both experiments, participants were presented with 6 temporal intervals ranging from 600 to 975ms in steps of 75ms and were asked to reproduce these intervals. In Experiment 1, which replicates Experiment 2 from Acerbi et al. (2012), the intervals were drawn from either a uniform or a skewed distribution in different experimental sessions (Figure 4.2A). Blocks of trials within each experimental session were divided into 6 randomized sequences, each of 96 trials, with each sequence reflecting the characteristics of the stimulus distribution for that session. In the skewed distribution, the 675ms interval was presented with a probability of 56/96, and each of the remaining stimuli with a probability of 8/96. Here we predicted that the standard deviation of the responses would be smaller in the skewed compared to the uniform distribution - responses would be more varied under the uniform distribution - reflecting the difference in the variance of the presented distributions: ~128ms in the empirical uniform distribution and ~107ms in the empirical skewed distribution (Prediction 1, Figure 4.3A). We also predicted that the overall mean response averaged across

durations and participants will be shorter in the skewed than the uniform distribution, reflecting the bias in the presented distribution of intervals toward shorter values – 787.50ms in the uniform condition, 731.25ms in the skewed condition (Prediction 2, Figure 4.3A).

In Experiment 2, participants were presented with time intervals drawn either from a uniform or a Gaussian distribution in different experimental sessions (Figure 4.2B). The empirical uniform distribution had a mean of 787.5ms and a standard deviation of ~128ms. The empirical Gaussian distribution had the same mean as the uniform distribution (787.5ms) and a standard deviation of ~93ms. Again, each block in a session consisted of 6 runs of 96 trials reflecting the statistical properties of the distribution. In the Gaussian distribution, the stimulus probabilities, rounded to the nearest whole number, were: 5/96 for 600ms and 975ms, 16/96 for 675ms and 900ms, and 27/96 for 750ms and 825ms. These parameters were chosen such that the Gaussian distribution is maximally distinguishable from the uniform distribution while at the same time not approaching a point distribution, as exposure to point distributions is known to produce repulsive aftereffects rather than regression towards the mean (e.g., Roseboom, Linares, & Nishida, 2015; Webster, 2015; Linares, Cos, & Roseboom, 2016). Repeated presentation of the same stimulus (adaptation) causes the perception of subsequently presented stimuli to be repelled from the original stimulus. In the domain of time perception, it has been found that adaptation to long intervals makes subsequent stimuli appear shorter, while adaptation to short intervals makes subsequent stimuli appear longer (Heron et al., 2011). In addition, we decided that a minimum of ~100 repetitions per stimulus would be needed to reliably measure performance, and if the Gaussian distribution were to be more peaked, we would not be able to collect enough data for the durations at the extremes of the range. Finally, we aimed to keep the range of the Gaussian and uniform distribution equal, and if the variance of the Gaussian were too narrow, this would also change the range of the distribution as the rounded probabilities of extreme values will drop to zero. Here we expected that participants' responses would be more biased towards the mean in the Gaussian distribution, and therefore predicted that the standard deviation of their responses would be smaller under the Gaussian than under the uniform distribution (Prediction 3, Figure 4.3B).

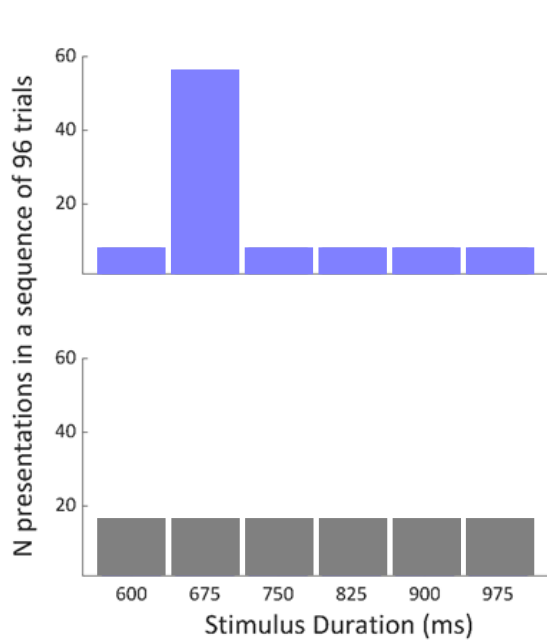
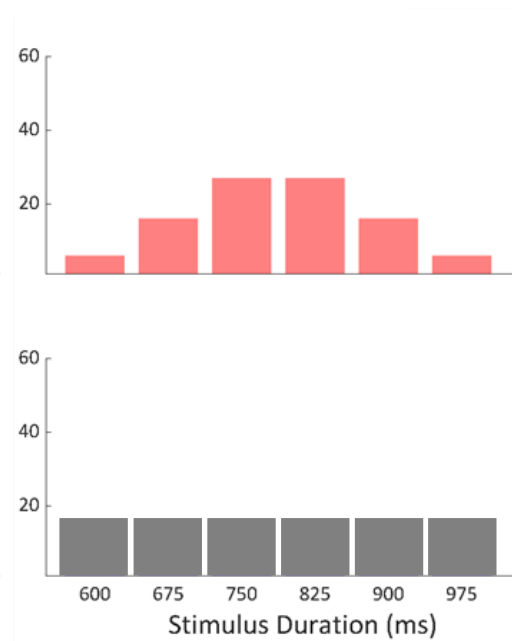
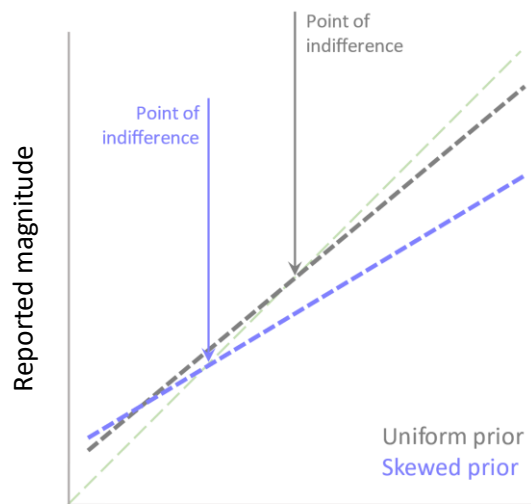
A. Experiment 1**B. Experiment 2**

Figure 4.2. Distributions of intervals presented in each condition. (A) In Experiment 1, participants were presented with intervals drawn from a skewed distribution (top) and a uniform distribution (bottom) in different experimental sessions. (B) In Experiment 2, participants were presented with a Gaussian distribution (top) and a uniform distribution (bottom) in different sessions.

A. Experiment 1**B. Experiment 2**

Stimulus magnitude

Figure 4.3. Predicted effects of prior shape on regression to the mean. (A) In Experiment 1, we predict that the magnitude of the bias towards the prior will be greater in the skewed compared to uniform condition. As a result, responses in the skewed distribution will have a smaller standard deviation than responses in the uniform distribution. We also predict that participants will be biased towards a shorter duration in the skewed than in the uniform distribution. (B) In Experiment 2, we predict that regression to the mean will be stronger in the Gaussian compared to the uniform distribution. As a result, responses in the Gaussian distribution will have a smaller standard deviation than responses in the uniform distribution. We also predict that in the Gaussian distribution participants will be biased towards a shorter duration than in the uniform distribution, but this difference will be smaller than the difference in Experiment 1.

In addition, in Experiment 2, the uniform and Gaussian distributions have the same mean, so we expected that participants' responses would regress towards the same value. However, duration estimation exhibits scalar variability (Weber's law, generally) – as durations increase, the variability associated with behavioural responses increases linearly in physical values (Gibbon, 1977; Church & Meck, 2003; Wearden, 2003; Lejeune & Wearden, 2006; Rammsayer, 2014). Under our Bayesian assumptions, this asymmetry in estimation variability would cause the absolute magnitude of underestimation of long durations to be greater than the overestimation of short durations. As already noted, there were fewer trials of the shortest and longest durations in the Gaussian compared to the uniform distribution, resulting in stronger overall regression to the mean effects for the Gaussian condition (see Prediction 3). The combination of asymmetry in regression effects introduced by scalar variability and stronger regression to the mean effects in the Gaussian condition overall should result in a small difference in the mean estimated duration between conditions such that durations in the Gaussian condition would be estimated as shorter than in the uniform condition (Prediction 4, Figure 4.3B).

The two conditions in each experiment were tested in separate experimental sessions. Sessions were completed on different days to minimize carry-over effects between conditions. The order of the conditions in the two experiments was counterbalanced between participants. If a participant completed both experiments, then the order of the conditions across the two experiments was pseudorandomized with the additional constraint that the skewed and Gaussian distributions would not be presented in successive experimental sessions. Sessions were completed at the same time of day to keep participants' levels of alertness similar across sessions (therefore, the minimum period between two sessions was ~24 hours). In addition to the main experimental sessions, participants completed two additional tasks designed to estimate perceptual and motor noise (see Procedure). Each task was completed in a different session.

4.2.3 Procedure

In both experiments, the main experimental task was a duration reproduction task following Acerbi et al.'s (2012) procedure (Figure 4.4). Each trial started with the presentation of a fixation cross, which was a cue that participants could initiate the interval-to-be-replicated. At their own pace, participants pressed a button on a button box, and after a time interval between 600 and 975ms (see Figure 4.4), a Gabor appeared on the screen for ~18.8ms (approximated by 3 consecutive frames; see Materials and Apparatus for characteristics of the

Gabor). 250ms after the Gabor disappeared, the fixation cross appeared again to prompt participants to reproduce the duration of the presented interval (the interval between initial button-press and Gabor disappearance). Temporal reproduction was performed by pressing and holding the button on a button box for the entire apparent duration of the interval. After a period between 300-500ms following a response, participants were presented with visual feedback for ~100ms (approximated by 16 consecutive frames) which was computed as the signed difference between reported and actual duration, $f(x, r) = r - x$. The colour of the feedback further signified whether participants were within 10% of the duration presented on the current trial. The feedback was bright green if the given response was within 10% of the presented duration, and brown if the error between the response and the presented duration was greater than 10%. Between 100-300ms after feedback presentation, the fixation cross disappeared and 100-300ms later the next trial began. Bayesian ideal-observer models sometimes include a parameter fitted to account for errors resulting from lapses in attention (Stengård & van den Berg, 2019; Acerbi et al., 2014). To reduce the effect of lapses without including free parameters in our model, on every trial, after the presentation of the stimulus, participants were given the option to opt-out instead of reproducing the presented duration. Participants were instructed to opt-out if they were not paying attention during the presentation of the interval on a given trial – they were distracted when pressing the button to begin the interval or missed the presentation of the Gabor which signalled the end of the interval. They were further instructed that they should not opt-out if it was simply a difficult trial, as terminated trials would be repeated later in the run, so there is no benefit to this strategy. Participants were not able to opt out after providing a response.

Participants completed multiple blocks of the duration reproduction task per session, depending on their performance. Each block consisted of 6 runs of 96 trials which reflected the statistical properties of the distribution for that session (see Figure 4.2). To monitor performance, we calculated Spearman's correlation coefficient between presented and reported time at the end of each block, and the fractional change in mean squared response error between every two successive blocks. Fractional change was calculated as the difference in mean squared error between two blocks divided by the mean squared error of the first of the two blocks. Following Acerbi et al. (2012), testing in Experiment 1 was terminated when participants' performance stabilized – the fractional change between two successive blocks must be less than 0.08. In the analysis, we included data only from these last two blocks in order to reduce the effects of learning on the results. In Experiment 2, we terminated testing when participants complete 3 blocks with stable reproduction performance. Here there was no

requirement for the three blocks to occur in immediate succession. This difference between the two experiments was decided because under the Gaussian distribution, the stimuli with lowest presentation frequency (600 and 975ms) were presented only 30 times in a block, and 2 blocks would not provide enough trials for analysis. Finally, trials were excluded and repeated if: (1) participants gave multiple responses, (2) failed to begin responding within 1500ms of the presentation of the response prompt, (3) responded before the appearance of the response prompt (4) gave responses outside of the window between 300ms and 1462ms (extending from 50% below the shortest duration to 50% above the longest duration in the range), or (5) decided to opt-out instead of reproducing the presented interval. After the end of each block, the recorded button presses were further examined for invalid responses. This included trials in which, due to issues with the button box, the recorded responses did not contain a clear marker for the start or the end of the response. These trials were excluded from the analysis.

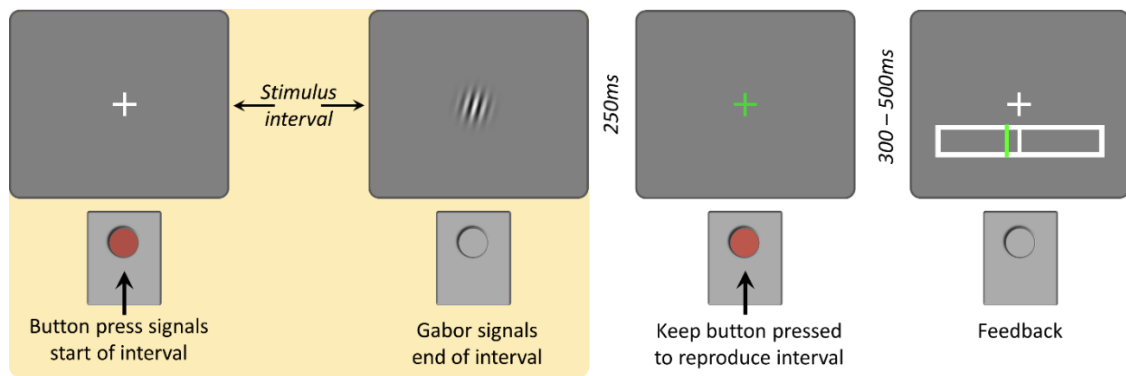


Figure 4.4. Duration-reproduction task. Participants' task was to reproduce the duration of a presented time interval. The interval began when participants pressed a button on a button box and ended when a Gabor appeared on the screen. Participants were asked to reproduce the interval by keeping the same button pressed for the same apparent duration as the interval. Finally, they received visual feedback regarding the accuracy of their response. In the feedback (the rightmost box) the white line in the middle of the box denotes perfect reproduction, while the coloured line denotes the error between participants' reproduced duration and the actual duration of the interval. The distance between the coloured and white line, therefore, denotes the magnitude of the error. The position of the coloured line relative to the white line denotes the direction of the error – if the interval was underestimated, the coloured line is to the left, and it is to the right if overestimated. The colour of the coloured line further denotes whether participants' response is within 10% of the duration presented on that trial (green) or not (brown).

One of the goals of these experiments was to build Bayesian ideal observer models to describe participants' behavioural reports. Two parameters in these models that we expected would vary between participants are (i) perceptual noise and (ii) motor noise. Therefore, in addition to the experimental sessions outlined above, participants completed two tasks designed to estimate these parameters. Originally, we planned for participants to complete both of these tasks after the main experimental sessions. However, due to the COVID-19 pandemic, some

participants who took part in both Experiment 1 and Experiment 2 completed the extra tasks in between the reproduction sessions of the two experiments.

To measure perceptual noise, we used a two-interval forced-choice task in which, on each trial, participants were presented with two intervals - a standard interval and a comparison interval - and indicated which of the two was apparently shorter in duration. The beginning of each interval was defined by a button press initiated by the participant, and the end of the interval by the presentation of a Gabor (~18.8ms; see Materials and Apparatus for characteristics of the Gabor). Participants completed 12 blocks (50-80 trials per block), and in each block, the standard interval was one of the six durations used in the main experiment (600, 675, 750, 825, 900, or 975ms). We used the Bayesian adaptive psychometric method QUEST (Watson & Pelli, 1983) to estimate the comparison duration at which participants correctly identified the shorter duration in a trial with 75% accuracy. On each trial, QUEST provided a suggestion of the optimal difference between the main and comparison stimuli that should be tested next. On trials where QUEST proposed a difference smaller than 0ms, the presented difference between the main and comparison stimuli was drawn from a uniform distribution spanning 0 to 10ms. Participants received feedback indicating whether their response was correct on each trial. We used the estimated duration threshold at each duration level as an estimate of participant-specific perceptual noise.

Initially, participants completed 50 trials after which we evaluated whether the estimate of the threshold provided by QUEST has converged. If at the end of a block the last 10 trials contained 5 consecutive trials between which the difference in the stimulus values suggested by QUEST was less than 0.5ms, we concluded that the estimate of the threshold has converged, and the block ended. If the difference between the suggested values was larger, we concluded that the threshold has not converged, and participants completed 10 more trials after which the convergence was evaluated again. The block ended either when the criterion for convergence was fulfilled or when participants reached 80 trials irrespective of whether QUEST has converged or not. The difference suggested by QUEST after the last trial in each block was taken as the estimate of perceptual noise. Plots showing the convergence of threshold for each participant and each stimulus are included in Appendix A (Figure A1). Overall, for all participants QUEST converged in the majority of blocks, and in at least one of the two blocks for each main stimulus (Appendix A, Table A1). While testing Participant 3, a hardware issue caused one of the blocks containing the 900ms stimulus and one of the blocks containing the 975ms stimulus to end too early. Participants' average thresholds for each stimulus level have also been plotted in Appendix A (Figure A2).

In addition to perceptual noise, we estimated motor noise using a reaction time task. In this task, at the beginning of each trial participants pressed a button that triggered the presentation of a low-contrast Gabor (see Materials and Apparatus for characteristics of the Gabor). Participants' task was to keep the button depressed while the Gabor was present and release the button as quickly as possible when the Gabor disappeared. The six stimulus durations from the main experimental task were presented in separate blocks. Each block consisted of 100 trials. Previous studies have reported that for healthy adults, the average simple reaction times to visual stimuli range between 150 to 260ms (Elliott, 1968; Wilkinson & Allison, 1989; Eckner, Kutcher, & Richardson, 2010). Therefore, if participants released the button before the disappearance of the Gabor or less than 100ms after it disappeared, the trial was discarded and repeated, as reaction times quicker than 100ms are likely to reflect stimulus anticipation. We fit an Exponentially Modified Gaussian (exGaussian) distribution to the reaction times to each stimulus using the DISTRIB toolbox for Matlab (Lacouture & Cousineau, 2008) and used the estimated standard deviation of the distribution as an estimate of motor noise.

Across participants, Experiment 1 took ~8 hours to complete and Experiment 2 took ~12 hours. Instructions for all tasks can be found in Appendix B.

4.2.4 Materials and apparatus

The experiment was run on Matlab2018b using Psychtoolbox-3.0.14 (Brainard, 1997, Pelli, 1997; Kleiner, Brainard, Pelli 2007) and presented on a colour-calibrated 22-in Iiyama Master Pro 513 CRT monitor with 160Hz refresh rate, and 800x600 pixel resolution. The experiment was presented on a grey background. Participants sat ~57cm away from the screen. The main stimulus was a black-and-white Gabor with 3dva size, 2c/d spatial frequency. The standard deviation of the Gaussian envelop of the Gabor was 0.33dva. The rotation angle and phase of the grating was randomized between 0 and 180° on every presentation to mitigate visual adaptation effects.

Participants provided responses through a custom-made button box that was connected to a professional grade sound card (M-Audio Delta 66) with a sampling rate of 44100 Hz. This allowed us to achieve an average error in measured latency of ~0.002ms for recording the timing between button presses and releases, and ~0ms (SD = ~1ms) for the timing between button press and visual outcome.

4.2.5 Participant exclusion criteria

We applied strict exclusion criteria to participants' performance to ensure the quality of the collected data. To ensure that participants were familiar with the structure of the duration-reproduction task and could read the feedback display, each session started with a practice block. In the practice block, participants were presented with a sequence consisting of 10 repetitions of a short, medium, and long stimulus presented in order – 400ms, the mean of the distribution used in the testing blocks, and 1175ms. The sequence was repeated twice. At the end of every trial in the practice block, the feedback remained on the screen until the participant pressed a button on the keyboard. The first testing block in each session was also designated for practice, and it was not included in the analysis.

In the main experimental task, any session was terminated, and participants were excluded from further participation in the study if they did not show sufficiently good task performance: (1) on the second testing block in each session, the Spearman correlation coefficient between presented and reported time had to be at least 0.60 and the slope at least 0.5 (an increase in participants' responses of at least 1 ms for every 2 ms increase in stimulus duration); (2) on the third testing block, the correlation between presented and reported time had to be at least 0.65 and the slope at least 0.5. Successful participants then completed additional blocks until they produced the necessary number of blocks needed for analysis. Blocks included in the analysis of each experiment had to exceed the performance thresholds above (Spearman correlation coefficient ≥ 0.65 and slope ≥ 0.5). In addition, like Acerbi et al. (2012), we wanted to analyse only data collected after participants' reproduction performance had stabilised to avoid measuring general task learning rather than the influence of acquired distributional statistics. Therefore, in Experiment 1 the absolute fractional change in mean squared error between the two last blocks had to be smaller than 0.08. In Experiment 2 where we required 3 blocks, there was no requirement for the blocks used in the analysis to be successive.

The original version of the experiment detailed in the stage 1 registered report included less training and more constricting exclusion criteria that were put in place to save time (for example, participants were to be excluded if their performance did not plateau within a certain number of blocks). However, participants found the task difficult, so we increased the amount of training. In addition, we also found that even with extra training, most participants were still finding the task difficult, so it was counterproductive to exclude capable participants who needed to complete more blocks to reach stable performance.

4.2.6 Data analysis

Bayesian ideal observer model

We followed Acerbi et al.'s (2012) approach to build Bayesian ideal observer models to predict each participant's responses to the stimuli in the two experiments. The model we used here is detailed in Chapter 3.1 (*Model 1*, p.46). Some criticisms of computational models of perception relate to the fact that during the analysis stage, researchers have multiple degrees of freedom, which ensures that any pattern of responses could be explained by some model that is derived post-hoc (Bowers & Davis, 2012). The present study used a fully deterministic model that left us no flexibility for fitting any model parameters. This allowed us to derive specific, falsifiable predictions regarding participants' behaviour that we tested using Bayesian statistics. The values for the two parameters in the model that were free to vary, sensory noise (w_s) and motor noise (w_m), were estimated empirically. For each participant and every duration, w_s was estimated through the previously described 2IFC task as the threshold at which participants could indicate the shorter of two stimuli with 75% accuracy. w_m was estimated as the standard deviation of the exGaussian distribution fitted to each participants' reaction times associated with each stimulus level. It is also important to note that the model used here assumes a fixed prior. It was not necessary to use a prior that updates across trials because the analysis was based only on data collected after participants reach stable performance, at which point we assumed that they would have internalized the properties of the stimulus distributions.

Statistical analysis

We conducted Bayesian t-tests to examine whether the skewed and Gaussian conditions give rise to responses with smaller variability than the uniform conditions (Predictions 1 and 3) and whether mean responses are shorter in the skewed and Gaussian conditions compared to the uniform conditions (Predictions 2 and 4). To test Predictions 1 and 3 we calculated the difference in the standard deviation of responses across all durations between distributions (within each condition we will calculate the standard deviation of all estimates, ignoring duration levels). We chose this measure because we wanted a single, simple metric for each condition that would be sensitive to the magnitude of the regression to the mean – when regression to the mean is stronger, participants' responses will be less varied (and vice versa). To test Predictions 2 and 4 we calculated the difference in mean responses across all durations between the relevant distributions, again averaged across participants. In all four Bayesian t-tests, the null hypothesis was defined as a point distribution at 0 difference, and the alternative hypothesis was defined as a half-normal distribution centred at 0 with a standard

deviation equal to the difference in distributions we expect to find. Each specific alternative hypothesis is detailed below. As our required sample size was below 30 participants (see section Participants), we applied a correction to standard error estimates as suggested by Dienes (2014) and Berry (1996) – standard errors were increased by a factor of $\left(1 + \frac{20}{df^2}\right)$, where df stands for degrees of freedom. Bayes factors (BF_{10}) larger than 3 were interpreted as providing evidence for the alternative hypothesis, and BFs smaller than 1/3 were interpreted as providing evidence for the null hypothesis (Dienes, 2014).

The expected difference between the standard deviation of responses under the uniform and skewed distributions (Prediction 1) and the uniform and Gaussian distributions (Prediction 3) were based on the Bayesian ideal observer model. For each participant, we calculated the difference in the standard deviation of the model-generated responses (across all stimulus durations) between the two distributions. Then, we found the average difference across all participant-based model estimates and used this as the standard deviation of the alternative hypothesis (H1).

For the comparisons of mean responses under the different distributions in each experiment (Prediction 2 and Prediction 4), we relied on the results reported in Acerbi et al. (2012). The authors report a one-sample frequentist t-test which revealed that the difference in response bias (the difference between actual and reported duration) between the skewed and uniform distribution was significantly different from zero ($M_{\text{difference}}(\text{uniform-skewed}) = 32.2\text{ms}$, $SD = 7.9\text{ms}$). Based on this result, we defined our expected difference in mean duration estimates (the standard deviation of H1) as 32.2ms.

To calculate the standard deviation and mean of behavioural responses for each condition, we sub-sampled trials from duration levels with greater presentation frequencies. To ensure that our results were not due to any specific subsample, we generated 1000 samples per participant and calculated the differences in mean and standard deviation. We then used the mean of the re-sampled distributions as the participant-specific estimates of the differences between conditions, which gave us a strong indication of the true parameters.

4.2.7 Required sample size

We estimated our required sample size using the Bayesian analysis outlined above on data simulated with Acerbi et al.'s (2012) Bayesian ideal observer model (see Dienes, 2014). We ran simulated experiments with different sample sizes (between $N=12$ and $N=20$) and checked what Bayes factor we expect to obtain with that sample size. To estimate the expected Bayes

factor for the difference in mean response between the different distributions in each experiment (Prediction 2 and Prediction 4), we based our estimate of prior effect size (the standard deviation of H1) on the difference reported by Acerbi et al. (2012): 32.2ms. To estimate the expected difference in the standard deviation of responses between stimulus distributions in each experiment (Prediction 1 and Prediction 3), we analysed Acerbi et al.'s (2012) raw data. For each participant (N=6), we extracted the data collected in the last 12 runs (~1000 trials), which would have corresponded to the data used in the original analysis. Next, we calculated the standard deviation of responses across all durations separately for the uniform and skewed distribution for each participant. Finally, we averaged the difference in standard deviation between the two distributions (uniform SD – skewed SD) across all participants to obtain our prior effect sizes for Prediction 1 and Prediction 3, which we found to be equal to 16.69ms. Finally, for comparisons that provided evidence for the alternative, we also investigated whether we would be able to conclude in favour of the null if we find no difference between the distributions. For these tests, the predicted effects were again set to the differences in Acerbi et al.'s (2012) data, and the measured effects were set to 0.

To mirror the design of the experiments, the simulated sequences for the comparisons between the uniform and skewed distributions consisted of 1152 trials (2 blocks of 576 trials), and the sequences for the comparisons between the uniform and Gaussian distributions consisted of 1728 trials (3 blocks of 576 trials). We also report the robustness region (RR) for each Bayes factor, which is the range of prior effect sizes which would lead to the same inferential conclusion. For example, if the reported Bayes factor supports the alternative hypothesis, the robustness region gives the minimum and maximum prior effect sizes which would also lead to Bayes factors that support the alternative hypothesis.

We varied the perceptual and motor noise across simulations to mimic individual differences in performance. Perceptual noise, defined as the just noticeable difference in duration discrimination tasks, increases proportionally with increase in the duration of the standard stimulus (e.g., Shi et al., 2013). Therefore, in our simulations, perceptual noise was defined as a scalar. Estimates of perceptual noise (Weber fraction) were sampled from a lognormal distribution centred at 0.157 (SD = 0.02) – the mean of the fitted perceptual noise parameters reported by Acerbi et al. (2012). The standard deviation of the sampling distribution was greater than the standard deviation reported in Acerbi et al. (2012) in order to simulate greater inter-participant variability. The generated estimates of perceptual noise varied between ~10-25% of the presented duration, which is in accordance with results reported by studies using 2IFC visual duration tasks to estimate the Weber fraction between 600-1000ms

(Rammsayer & Pichelmann, 2018; Rammsayer, 2014; Lapid, Ulrich, & Rammsayer, 2008).

When analysing the behavioural data from the experiment, the perceptual noise parameter for every participant was directly derived from their responses on the 2IFC task. To our knowledge, no studies to date have examined whether motor variability remains constant or increases proportionally to stimulus duration. To evaluate both of these possibilities, we defined motor noise as constant in one set of simulations, and as proportional to stimulus duration in another set of simulations. Estimates for scalar motor noise were sampled from a lognormal distribution with a mean of 0.072 (SD = 0.01) – the mean of the fitted motor noise parameters reported by Acerbi et al. (2012). Again, we increased the standard deviation of the sampling distribution to avoid underestimating inter-participant variability. The generated estimates of scalar motor noise varied between ~13-20% of each stimulus duration. In the simulations where motor noise was constant across durations, the estimates of scalar noise were multiplied by a constant (787.5), generating parameters that varied between ~100-170ms. In the analysis of the experimental data, the motor noise parameter for each participant was derived directly from their performance on the reaction time task.

In our initial simulations, the model provided realistic values for the mean differences across conditions but underestimated the variability we can expect to find across participants in the behavioural experiment (see Table 1), which leads to unrealistically high estimates of expected effect size. This is likely because the model does not account for random variability across participants, such as fatigue and motivation, which will affect the behavioural results. Therefore, to more accurately estimate our expected effect size, we added a parameter to simulate measurement noise to the difference between the compared conditions for each simulated participant. The corrupted difference scores were randomly chosen from Gaussian distributions centred at the ideal differences produced by the model with standard deviations that were varied between 5-15ms across the simulated participants. We ran 1000 simulations for each combination of sensory, motor, and measurement noise. It is important to note that measurement noise is not part of the Bayesian ideal observer specified above, and this was not included in the analysis of the experimental data – we added additional noise to the simulations solely to provide more realistic results to help determine desired sample size.

Difference in standard deviation of simulated responses (N=12)

	Constant motor noise		Scalar motor noise	
	No added noise	With added noise	No added noise	With added noise
Prediction 1: Mean Uniform – Skewed (SE)	13.78 (0.20)	13.63 (3.43)	14.05 (0.21)	14.05 (3.78)
Prediction 3: Mean Uniform – Gaussian (SE)	15.07 (0.26)	15.27 (3.49)	15.09 (0.24)	14.77 (3.40)

Difference in mean of simulated responses (N=12)

	Constant motor noise		Scalar motor noise	
	No added noise	With added noise	No added noise	With added noise
Prediction 2: Mean Uniform – Skewed (SE)	26.79 (1.57)	26.74 (3.69)	27.32 (1.36)	26.34 (3.83)
Prediction 4: Mean Uniform – Gaussian (SE)	0.48 (0.21)	0.22 (3.48)	0.12 (0.26)	0.72 (3.40)

Table 4.1. Results from simulations. Average difference in mean and standard deviation of the simulated responses across the two experiments with and without added measurement noise.

The simulations with added measurement noise revealed that with a sample size of 12 we could expect to exceed our sensitivity threshold for the comparisons of standard deviations across distributions (Prediction 1 and Prediction 3). For both of these predictions, we specified a prior effect size of 16.69ms. Regarding Prediction 1, the simulations with constant motor noise showed that the average standard deviation was higher under the uniform distribution compared to the skewed distribution, $M_{\text{difference}} = 13.63\text{ms}$ ($SE = 3.43$) which provided support for the alternative hypothesis with $BF_{10} = 772.66$, $RR [1, 1000]$. Alternatively, when we specified a measured difference of 0ms ($SE = 3.43$), the data provided evidence in favour of the null hypothesis, $BF_{10} = 0.20$, $RR [9.7, 1000]$. Regarding Prediction 3, the simulations with constant motor noise revealed that the average standard deviation was higher under the

uniform than the Gaussian distribution, $M_{\text{difference}} = 15.27\text{ms}$ ($SE = 3.49$), and we found evidence for the alternative hypothesis, $BF_{10} = 4003.47$, $RR [0.9, 1000]$. When the measured difference was set to 0ms ($SE = 3.49$), we found evidence in favour of the null hypothesis, $BF_{10} = 0.20$, $RR [9.8, 1000]$. We found similar results when motor noise was proportional to the stimulus duration (see Table 4.1).

For the comparisons of mean responses across distributions (Prediction 2 and Prediction 4), the simulations again revealed that we could expect to exceed our sensitivity threshold with a sample size of 12. The prior expected effect for both predictions was set to 32.2ms – the difference in mean responses reported by Acerbi et al. (2012). In the simulations in which motor noise was constant, the average difference in mean responses between the uniform and skewed distributions (Prediction 2) was 26.74ms ($SE = 3.69$). This result provides strong evidence for the alternative hypothesis with $BF_{10} = 4.26 \times 10^{10}$, $RR [0.55, >1000]$. When the measured difference was set to 0ms ($SE = 3.69$), we found evidence in favour of the null, $BF_{10} = 0.12$, $RR [10.40, 1000]$. The average difference in mean responses between the uniform and Gaussian distribution (Prediction 4) produced in the simulations was 0.22ms ($SE = 3.48$). In contrast to the results regarding Prediction 2, this provides strong evidence in favour of the null hypothesis ($BF_{10} = 0.12$, $RR [10.35, 1000]$). We found similar results when motor noise was proportional to the stimulus duration rather than constant (See Table 4.1).

The simulations detailed above revealed that we could expect to find moderate to very strong evidence either in favour of the null or the alternative hypothesis for a wide range of possible prior effects. This suggests that even if the model under- or overestimates the differences between conditions because we have misspecified participants' true noise parameters, we would still likely arrive at the same inferential result. We computed additional simulations to examine further how robust our analyses would be to the over- and underestimation of the perceptual and motor noise parameters. The results from these simulations have been included in Appendix C. In brief, the comparison between the standard deviation of the two distributions is relatively robust to parameter misspecification. The sensitivity threshold of $BF_{10} > 3$ was reached with a reasonable range of motor and perceptual noise parameters. The difference predicted by the model is too small only when perceptual noise is grossly underestimated, or when there is a large discrepancy between motor and perceptual noise – for example when perceptual noise is very high but motor noise is very low. We believe it is unlikely to observe such extreme deviations from the true estimates of participants' noise parameters. For more details, please refer to Appendix C.

4.2.8 Participants

Acerbi et al.'s (2012) experiment was completed by 6 participants. Based on the estimates of sensitivity for the proposed Bayesian t-tests discussed above, we aimed to recruit a starting sample of 12 adult participants (18 years of age or older) for each experiment and conduct our analyses. We were primarily interested in the comparison between the standard deviation of the distributions in the two experiments. Therefore, we planned that if our results for these tests do not exceed the set sensitivity criterion ($BF_{10} < 1/3$ or $BF_{10} > 3$), we would continue data collection and redo the analysis after every additional participant until we reach a maximum of 20 participants per experiment.

As of now, we have recruited 47 participants to take part in the study. Of them, 16 participants ($Mean_{Age} = 23.81$ years, $SD_{Age} = 4.55$) fulfilled the performance criteria, but 6 dropped out due to the pandemic. From the remaining participants, 6 participants completed both experiments, 1 participant completed only Experiment 1, and 2 participants completed only Experiment 2. Therefore, the results presented here are based on 7 participants in Experiment 1, and 8 participants in Experiment 2.

4.3 Results

4.3.1 Effect of variance on responses

The main aim of this study was to investigate the effect of distributional variance on regression to the mean, with our overall hypothesis being that distributions with a smaller variance will lead to greater regression towards the mean. In Experiment 1, participants were presented with durations drawn from a uniform or skewed distribution, while in Experiment 2 participants were presented with durations drawn from a uniform or Gaussian distribution. We predicted that participants' responses would show greater regression towards the mean, and therefore smaller standard deviation, under the skewed (Prediction 1) and Gaussian distributions (Prediction 3) compared to the uniform distribution.

First, we found that, on average, the standard deviation of participants' responses was higher under the uniform than the skewed distribution, mean difference = 1.85ms ($SE = 2.74$). The Bayesian ideal observer model predicted a mean difference of 11.29ms, and the behavioural results provided anecdotal evidence in support of the null, $BF_{10} = 0.43$. In Figure 4.5, we can see that the predictions of the model participants' data diverge for the longest durations in the range. The model predicted that, on average, under the skewed distribution, participants' responses to the 825, 900, and 975ms stimuli would be more biased towards the mean than

under the uniform distribution. However, in the behavioural data, average responses to the three longest stimuli did not differ substantially between the two conditions.

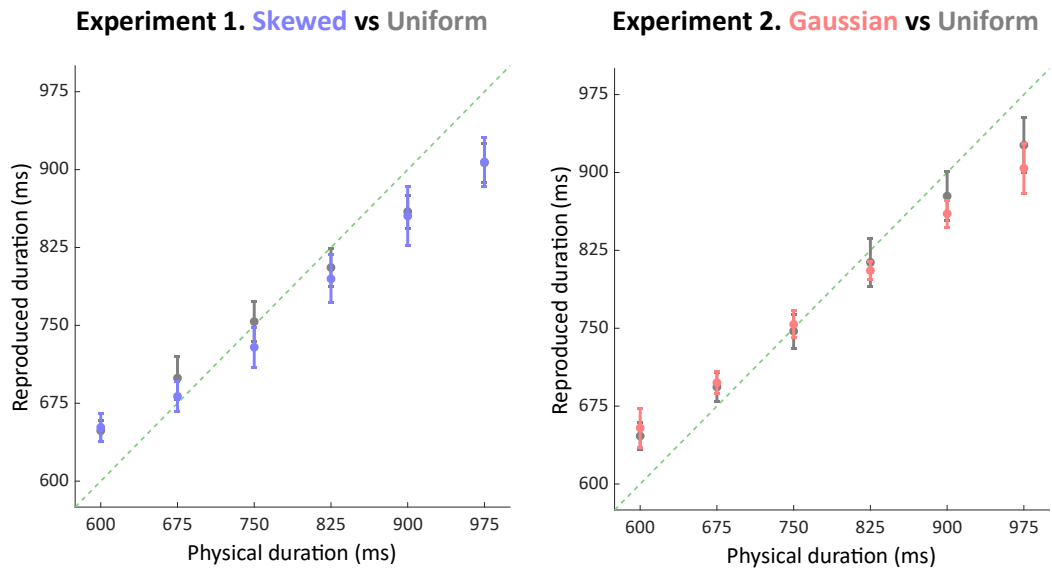
In support of Prediction 3, on average the standard deviation of participants' responses in Experiment 2 was higher under the uniform than the Gaussian distribution, mean difference = 19.12ms (SE = 9.41), and we found that these results provide moderate evidence in support of the effect predicted by the Bayesian ideal observer model, predicted difference = 15.34ms, $BF_{10} = 4.48$, RR [7.15 42.75]. Figure 4.5 also shows no substantial differences in the behavioural and model-predicted results. The pattern of averaged responses in the behavioural data is well-reflected in the average responses predicted by the Bayesian ideal observer model.

4.3.2 Effect of mean on responses

We predicted that participants' responses would be biased towards the mean of the underlying stimulus distribution. In Experiment 1, the mean of the skewed distribution was shorter than the mean of the uniform distribution, so we predicted that participants responses would be shorter in the skewed than the uniform distribution (Prediction 2). In Experiment 2, we also predicted that mean responses will be shorter in the Gaussian than the uniform condition due to greater regression towards the mean for longer compared to shorter durations (Prediction 4). For both predictions, we tested the difference in the behavioural data against the difference in mean responses between the uniform and skewed distribution reported by Acerbi et al. (2012) in their second experiment: 32.2 ms.

As predicted, in Experiment 1, mean responses were shorter in the skewed compared to the uniform condition, mean difference = 8.87 (SE = 4.99). However, the obtained difference was smaller than the prior expected effect and consequently provided only anecdotal evidence in favour of the alternative hypothesis, $BF_{10} = 1.37$. Finally, in Experiment 2 mean responses were 4.68ms (SE = 9.63) shorter in the Gaussian than uniform condition, but that difference was considerably smaller than the expected effect and provided anecdotal evidence in favour of the null hypothesis. $BF_{10} = 0.44$.

Behavioural results



Model predictions

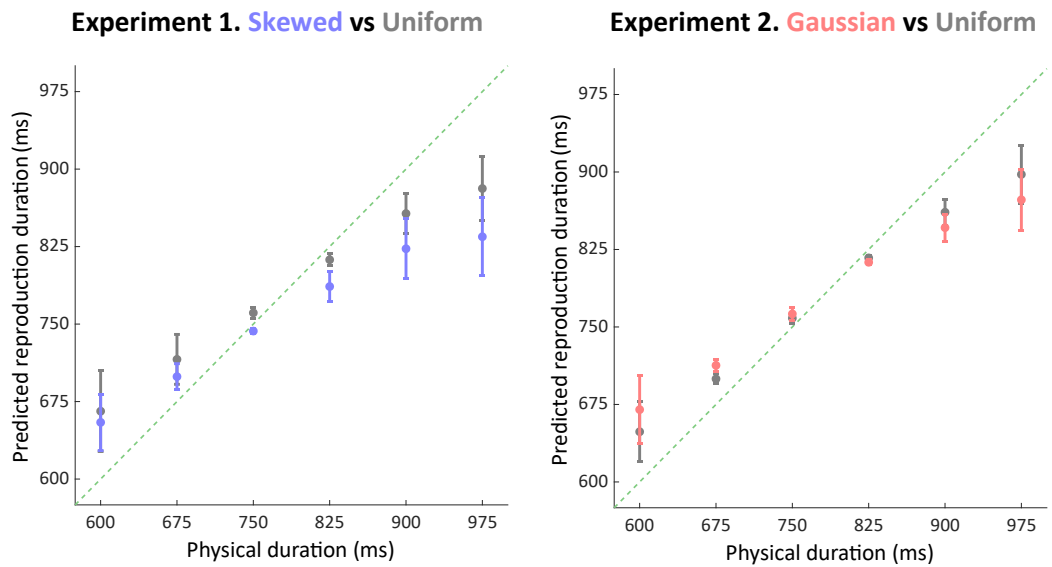


Figure 4.5. Behavioural results and model predictions for each experiment. The figures on the top row show responses to each stimulus duration averaged across participants, separately for the uniform and skewed distribution in Experiment 1 (left), and the uniform and Gaussian distribution in Experiment 2 (right). The bottom figure shows the average responses predicted by the Bayesian ideal-observer model for the conditions in each experiment. The green dotted line represents veridical responses. Error bars show $\pm 1SD$, calculated as the standard deviation of participants' average responses.

4.4 Discussion

In this study, we investigated whether the variance of the prior distribution influences the magnitude of perceptual biases towards the prior. Replicating and extending a study by Acerbi et al. (2012), we ran two experiments in which participants had to reproduce durations drawn from distributions with different variances, and implemented a Bayesian ideal-observer model to predict participants' regression to the mean effect in each condition. In Experiment 1, which replicated Acerbi et al. (2012), participants were presented with durations drawn from a skewed or uniform distribution, while in Experiment 2 durations were drawn from a Gaussian or uniform distribution. In Experiment 2 we found that, on average, participants' responses supported the prediction of the Bayesian ideal-observer model – responses were biased more towards the mean of the stimulus distribution under the Gaussian compared to the uniform distribution. In contrast, in Experiment 1, we found that participants' responses diverged from the prediction of the model. While the model predicted that participants would be more biased towards the mean in the skewed distribution, the behavioural data revealed a considerably smaller (almost non-existent) difference in the magnitude of regression to the mean than predicted by the model. Although we did not find conclusive evidence in favour of the null hypothesis, these preliminary findings suggest that the Bayesian model was not able to capture the pattern of participants' responses. In each experiment, we also investigated whether there was a difference in the value towards which participants regressed under the different stimulus distribution. In both experiments, we found that the mean of participants responses was shorter in the peaked than in the uniform conditions, and even though the results did not provide conclusive evidence for either the null or the alternative hypothesis, the differences we found were smaller than the result reported by Acerbi et al. (2012).

Previous studies investigating how the precision of the prior distribution affects responses confounded differences in variance with concurrent changes in range, making it impossible to divulge the specific effect of prior variance on perception (Berniker et al., 2010; Acerbi et al., 2014; Körding et al., 2004; Miyazaki et al., 2005; Garrido et al., 2013; Garrido et al., 2019). A fundamental prediction of Bayesian brain theory is that specifically the variance of the prior affects the magnitude of perceptual biases. Until now, this assumption had not been tested directly. In Experiment 2, we show that participants' responses reflect changes in the variance of the underlying distribution when keeping mean, range, and skewness constant. Studies in the field of ensemble perception have shown that participants can estimate the variance of stimulus distribution when explicitly asked to do so (e.g., Haberman et al., 2015; Solomon, Morgan et al., 2011; Suárez-Pinilla et al., 2018). Our results here further show that participants

also implicitly estimate the variance of stimuli presented over time to optimize future behaviour.

Contrary to our results in Experiment 2, we found that participants' responses in Experiment 1 did not support the prediction of the Bayesian ideal-observer model defined by Acerbi et al. (2012). While the model predicted that participants' responses would be biased towards the mean more in the skewed compared to the uniform distribution, there were very small differences between the two conditions in the behavioural data. The divergence between the model's predictions and the behavioural data is most striking for the three longest durations in the range – 825, 900, and 975ms. The model predicts that as the duration of the stimulus increases from 825 to 975ms, the regression to the mean in the skewed condition would increase relative to the uniform condition. In contrast, we see the opposite in participants' data – the difference between the two distributions is greatest at 750ms, beyond which the difference in responses between the two distributions progressively decreases, which is reminiscent of results reported by studies investigating adaptation effects (e.g., Roseboom, 2019). These results contradict the findings of Acerbi et al. (2012) who found that the Bayesian ideal observer model provided a good fit to participants' responses in the skewed condition. Interestingly, the researchers also found that the Bayesian ideal observer model could explain behavioural responses obtained in a highly skewed distribution where the overrepresented stimulus appeared on ~80% of the trials, compared to ~60% of the trials in their "standard" skewed distribution which we also used here. As the samples recruited in these experiments, including ours, are relatively small, it is possible that the differences between our results and the results reported by Acerbi et al. (2012) could be explained by individual differences, as others have found wide variation in the spread and strength of attractive and repulsive biases across participants (Bliss, Sun, & D'Esposito, 2017).

The pattern of results in the skewed condition is reminiscent of results reported by studies investigating adaptation effects (e.g., Roseboom, 2019), and suggests that the skewed distribution could be giving rise to opposing biases in different parts of the stimulus distribution – the stimuli closest to overrepresented duration are biased towards it, while more distant durations are biased away from it. In their most general formulations, Bayesian ideal-observer models are not equipped with the mechanisms necessary to predict repulsive after-effects. In contrast, models focused on explaining adaptation effects (e.g., Heron et al., 2012) are not able to predict positive biases, and overall, there has been little attempt to bridge the gap between these distinct classes of models. However, as the results in Experiment 1 show, repulsive and attractive biases can co-exist, and a model capable of predicting only

one type of bias falls short of describing participants' behaviour. It was recently shown that a hierarchical model consisting of a sensory layer based on population codes (which gives rise to adaptation), and a decision layer based on Bayesian inference (which gives rise to biases towards the prior) was able to predict positive and negative biases in estimating motion variance (Suàrez-Pinilla, 2018). Such an approach has the potential to provide a unifying account of two seemingly conflicting perceptual phenomena, and thus allow researchers to explain complex patterns of behaviour that are otherwise not well captured by the individual models alone.

Chapter 5. Effect of sensory precision on behavioural and neuroimaging perceptual biases in duration reproduction and estimation tasks

In the previous chapter, I presented two experiments in which we investigated how manipulating the variance of the prior distribution influences regression to the mean. Conversely, Bayesian brain theories predict that the magnitude of perceptual biases also depends on the variance of the sensory likelihood, but the existing research investigating the influence of sensory precision on regression to the mean is limited. Similarly, little is known about the potential neural correlates of regression to the mean. Here I present two experiments in which we investigated how sensory precision influences the magnitude of behavioural and neuroimaging perceptual biases in duration perception tasks. In both experiments, participants were presented with durations drawn from two overlapping distributions – a short and a long distribution. In Experiment 1 (N=20) we found that participants' behavioural estimates of the duration of the overlapping stimulus differed between the two conditions, and using EEG (N=19) we found that a multivariate classifier was able to distinguish stimulus context on the basis of the brain activity associated only with the overlapping stimulus. However, we found evidence of no difference in the magnitude of the behavioural or neuroimaging bias towards the prior between participants with low and high sensory precision. Furthermore, our preliminary results from Experiment 2 (N=27) reveal no conclusive evidence in favour of a correlation between regression to the mean and sensory precision under conditions of low motor noise. Altogether our results suggest that either the process that leads to regression to the mean-type effects might not be entirely Bayesian in nature or that the measures researchers use to estimate constructs like sensory precision and regression to the mean need to be properly validated.

5.1 Introduction

A key prediction of Bayesian models is that the resulting probability distribution, the posterior, will be biased more or less towards the prior depending on the relative variability of the likelihood relative to the variability of the prior – as the variability of the sensory likelihood decreases, the bias towards the prior will also decrease (see Figure 4.1 for an illustration of the effect of the relative variance of the likelihood and prior on the posterior). Here we are investigating how the precision of the sensory likelihood influences the magnitude of

perceptual biases in time perception using computational modelling and electroencephalography, where sensory precision is defined as the inverse of the likelihood variance (i.e., the greater the variance of the likelihood, the smaller the precision and vice versa).

Following the assumption of Gaussianity, an observer's best prediction before the presentation of new sensory information will be equivalent to the mean of the prior distribution. As already discussed in previous chapters, in magnitude estimation tasks (wherein participants must directly estimate the value of the presented stimulus), perceptual biases towards the prior can be measured based on the regression to the mean in observers' responses – stimulus magnitudes smaller than the mean tend to be overestimated, while stimulus magnitudes greater than the mean tend to be underestimated (Lejeune & Wearden, 2009; Jazayeri & Shadlen, 2010; Petzschner et al., 2015). The effect is range-specific such that the same physical stimulus will be over- or underestimated depending on its position in the prior distribution (Figure 5.1A). Therefore, based on the assumption of Bayesian models discussed above, we can predict that the precision of the sensory likelihood would influence the magnitude of regression to the mean – the less precise the sensory likelihood, the stronger the bias towards the prior. Alternatively, as sensory precision increases, regression to the mean will decrease, and participants' responses will be closer to the true stimulus magnitude (Figure 5.1B). This, in turn, means that the difference in reported magnitude between the same physical stimulus presented in two different contexts will decrease as sensory precision increases.

To date, there is very limited evidence investigating the relationship between sensory precision and perceptual biases in behaviour. Previous studies have found that there is a positive correlation between the regression to the mean participants exhibit in duration reproduction tasks and their sensory precision measured through a duration bisection task (Cicchini et al., 2012). In addition, studies have also found that musicians, and expert drummers, in particular, experience weaker regression to the mean than participants with no musical experience (Cicchini et al. 2012; Aagten-Murphy et al. 2014). Neuroimaging studies provide evidence that priors can also influence the brain activity associated with a specific sensory stimulus, suggesting that the effect of sensory precision predicted by Bayesian perceptual models may not be limited to behavioural data. Studies within the domain of time perception have found that the amplitude of the CNV is influenced by the duration presented on the previous trial (Wiener & Thompson, 2015; Damsma et al., 2020), and by the range of the distribution from which stimuli are drawn (Damsma et al., 2020). On a trial-to-trial level,

these studies showed that across a duration bisection (Wiener & Thompson, 2015) and a duration reproduction task (Damsma et al., 2020) the amplitude of the CNV on the current trial became more negative following longer durations presented on the previous trials. Damsma et al. (2020) also reported that in a duration reproduction task the amplitude of the CNV elicited by the same physical duration was more negative when the stimulus was embedded in a range of long compared to a range of short durations.

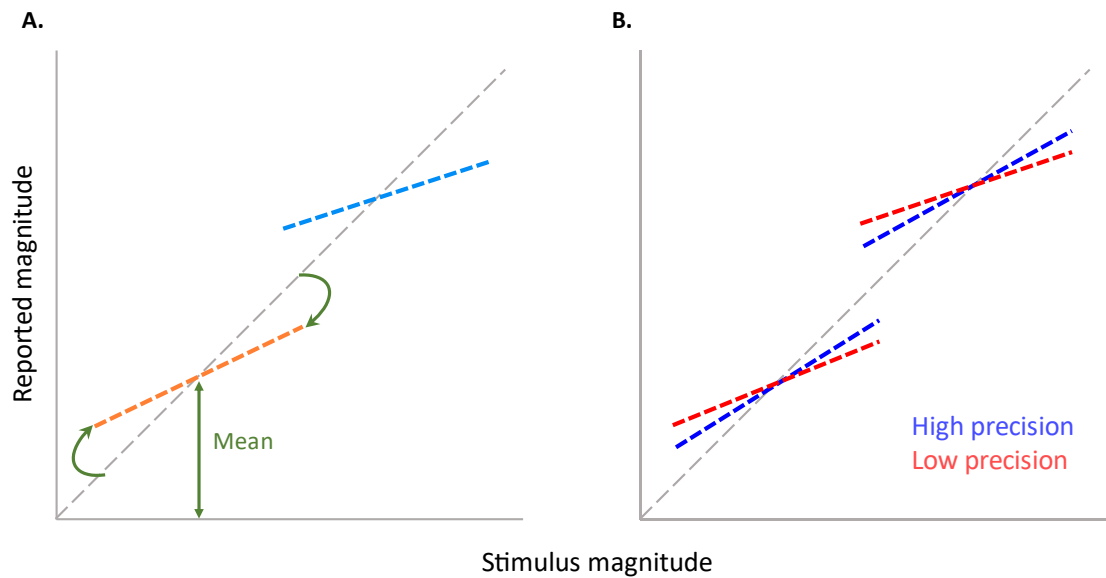


Figure 5.1. Regression to the mean. (A) Overall regression to the mean effect across two partially overlapping ranges – stimulus magnitudes greater than the mean are underestimates, while stimulus magnitudes smaller than the mean are overestimated. Due to scalar variability, regression to the mean is stronger for longer durations, so responses in the long range (in blue) are more biased to the mean than responses in the short range (in orange). **(B)** Effect of sensory precision on regression to the mean predicted by Bayesian brain models – high sensory precision (blue) results in weaker regression to the mean than low sensory precision (red). Responses of participants with low sensory precision are closer to veridical, and there is a smaller difference in responses to the overlapping stimulus. Again, due to scalar variability, regression to the mean is stronger in the long range than in the short range. In both figures the grey dotted line represents veridical responses.

Effects of stimulus context on EEG activity have been found not only in timing tasks but in other perceptual domains as well. Studies on pitch perception have found that the mismatch negativity (MMN) evoked response is affected by the variance and range of the underlying stimulus distribution – the smaller the variance and range of the prior, the greater the MMN amplitude elicited by the same physical stimulus (Garrido et al., 2013; Garrido et al., 2016; Randeniya et al., 2019). It has also been suggested that sensory precision affects the MMN response associated with a deviant stimulus, with greater sensory precision being related to an MMN response with higher amplitude. Kraus et al. (1996) showed that children with learning difficulties had worse syllable discrimination performance and smaller MMN response in

relation to syllable change compared to age-matched controls who could discriminate syllables with smaller acoustic differences and showed greater MMN response. These results suggest that the more different two stimuli are perceived to be, the more differentiable their associated brain activity will be.

As already discussed, in regression to the mean the same physical stimulus is perceived to be greater or smaller than the mean depending on the range of the stimulus distribution it is embedded in. In a preliminary analysis of the experiment presented here (Baykova & Roseboom, 2019) and a recent study by Damsma et al. (2020) it was shown that a classifier could distinguish stimulus context based solely on the EEG activity associated with a physical stimulus common to both distributions – the stimulus was the longest duration in one range of stimuli, and the shortest duration in the other range of stimuli. We can further predict that as the regression to the mean increases, the brain activity associated with a given stimulus will carry more information about the underlying stimulus distribution. Bayesian perceptual models predict that observers with low sensory precision will exhibit stronger regression towards the mean. Therefore, we can predict that in a Bayesian observer with low sensory precision, the brain activity elicited by a given stimulus will vary substantially depending on the overall stimulus distribution (Figure 5.2). Conversely, a Bayesian observer with high sensory precision will be less affected by context, and the brain activity associated with a given stimulus will differ less across different stimulus distributions. In other words, we assume that the brain activity generated by a given stimulus is a mixture of signals reflecting not only processing of the incoming physical information, but also signals related to the observer's subjective experience. If we assume that brain activity reflects observers' subjective experience, we can predict that in a Bayesian observer with low sensory precision, the brain activity elicited by a given stimulus will carry more information about stimulus context compared to a Bayesian observer with high sensory precision (who will experience a given stimulus identically irrespective of the underlying distribution, Figure 5.2).

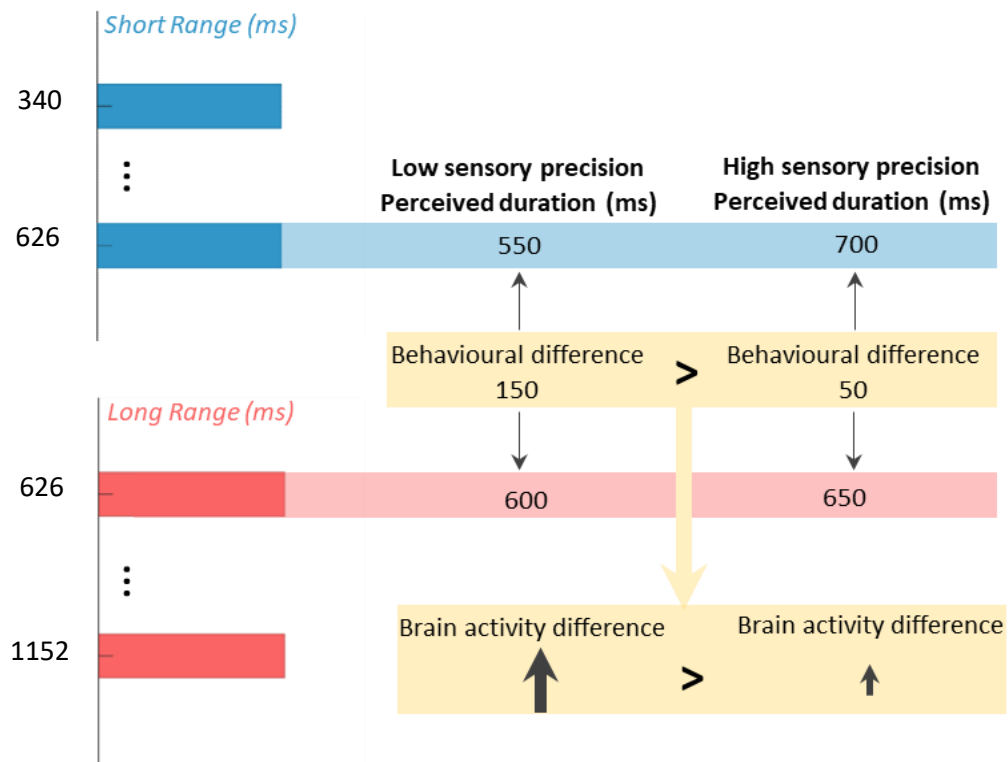


Figure 5.2. Effect of sensory precision on perceptual biases in brain activity. An observer with low sensory precision will experience the same physical stimulus, 626ms, very differently depending on the context the stimulus is presented in. This should result in a large difference in brain activity when the same stimulus is presented in different contexts. In contrast, the perceived duration of an observer with high sensory precision will be less affected by context. Therefore, the difference in brain activity when the same stimulus is presented in different contexts will be smaller.

Here we present two experiments in which we investigated the effect of sensory precision on perceptual biases across two duration perception tasks. In Experiment 1, participants completed a duration reproduction task in which they were presented with durations drawn from a short (340 to 626ms) or long range (626 to 1152ms). Our first hypothesis was that participants would exhibit regression towards the mean – they would overestimate the duration of the 626ms stimulus when it is presented in the context of the long range, but they would underestimate its duration when it is presented in the context of the short range. We will use a Bayesian ideal observer model to derive a specific, numerical prediction for this effect. Having established the presence of regression to mean in our dataset, we examined our main behavioural hypothesis – that observers with low sensory precision would exhibit stronger regression to the mean than observers with high sensory precision. Low and high sensory precision groups were derived using a median split. In Experiment 1, we also used EEG to investigate how sensory precision influences neuroimaging proxies of perceptual biases. In the EEG session participants were passively presented with the same stimuli they saw in the

behavioural experiment. Our prediction was that stimulus context would influence the brain activity associated with the same physical stimulus (626ms). Our first hypothesis was that an MVPA classifier would be able to distinguish stimulus context significantly better than chance when it is trained and tested only on the EEG data associated with the overlapping stimulus. In this case, we would be able to use classification performance as an indication of perceptual bias on the neuroimaging level, and our prediction was that context classification would be better in participants with low compared to high sensory precision. Finally, we also carried out univariate analysis examining how context influenced the amplitude of the CNV associated with the overlapping stimulus. Our prediction was that the CNV would have a greater amplitude (become more negative) when the overlapping stimulus is presented in the context of long compared to short durations. We also predicted that the difference in amplitude would be greater in participants with low compared to high sensory precision.

In Experiment 2, participants completed a direct duration estimation task in which they reported the duration of the presented intervals on a visual analogue scale. In duration-reproduction tasks, there is a potential discrepancy between perceived and reported duration due to confounding factors related to the method of providing responses (for example, motor noise). Reports in direct duration-estimation tasks are less likely to be affected by confounding factors unrelated to the sensory processing of the stimulus, and could therefore provide a more accurate representation of the relationship between sensory noise and regression towards the mean. Here we treated sensory precision as a continuous rather than categorical variable, and our prediction was that sensory precision and regression towards the mean would be positively correlated.

5.2 Experiment 1

5.2.1 Method

Pre-registration of behavioural experiment and data availability

The hypotheses, design, experimental procedure, and analysis plan for the behavioural experiment are pre-registered at the Open Science Framework (<https://osf.io/nbf38>). The behavioural data is available here: <https://osf.io/qgtfe/files/>. The EEG data is available here: <https://osf.io/u9kbv/files/> (this includes: raw EEG data, pre-processed EEG data, list of removed epochs, list of removed ICA components, results from the MVPA analysis, data for the ERP analysis).

Design

In this experiment, we investigated how sensory precision influences behavioural and neuroimaging perceptual biases towards the prior. The sensory precision of each participant was estimated based on the average 75% threshold reached in a 2-interval forced-choice (2IFC) duration discrimination task. We used a median split based on the average 75% threshold to divide participants into high and low sensory precision groups.

We quantified behavioural perceptual bias as the magnitude of regression to the mean (greater regression to the mean signifies greater bias) in a duration reproduction task. Participants were asked to reproduce durations drawn from a range of 5 stimuli between 340-626ms (short range), and a range of 5 stimuli between 626-1152ms (long range). Regression to the mean was defined as the difference in reproduced durations in response to the stimulus present in both conditions, 626ms (Figure 5.3). Notably, the durations follow a logarithmic scale to account for Weber's law – the variability of behavioural responses increases proportionally with increase in stimulus magnitude.

To investigate perceptual biases on the neural level, participants also completed an EEG session during which they were presented with the same ranges of durations. In the EEG task, we estimated perceptual bias based on the accuracy with which a multivariate classifier decodes stimulus context based on activity associated with the same physical stimulus – the 626ms (better classification performance signifies greater bias).

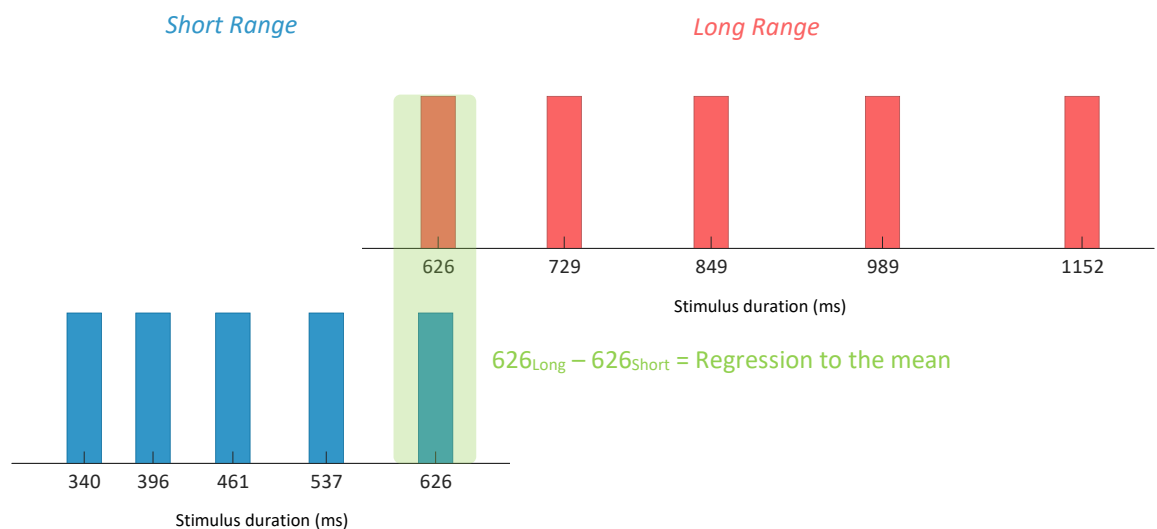


Figure 5.3. Stimulus conditions in the duration reproduction task. In different blocks participants were presented with stimulus durations drawn from a short or a long range. There was one stimulus common to the two ranges, 626ms, responses to which were used to quantify regression towards the mean.

Apparatus

The experiment was run on Matlab2018b using Psychtoolbox-3.0.14 (Brainard, 1997, Pelli, 1997; Kleiner, Brainard, Pelli 2007). The behavioural tasks were presented on a 22-in Iiyama Master Pro 513 CRT monitor, and the EEG task was presented on a 22-in LaCie Electron blue IV CRT monitor. Both monitors were gamma-corrected, had a resolution of 800x600 pixels, and a refresh rate of 120Hz. Participants sat ~57 cm away from the screen in all sessions. The main stimulus was a stationary cloud of 17 black and 17 white dots presented on a grey background. Each dot had a size of 0.4dva, and the overall size of the entire cloud was 12dva. The position of each dot was randomized on every trial to mitigate visual adaptation effects. In the behavioural sessions, participants provided responses on a gaming computer keyboard, and in the EEG session – on a standard keyboard.

Procedure

Participants completed 3 sessions in total. The first two sessions comprised of two behavioural tasks – a duration reproduction task, and a 2IFC task (completed in that order). The third session involved only passive observation of durations completed during EEG recording. Each of the behavioural sessions took ~2.5 hours to complete, and the EEG session took ~2 hours. Instructions for all tasks can be found in Appendix D.

Duration reproduction task

The main behavioural task in this experiment was a duration reproduction task (Figure 5.4). A white fixation cross was presented on the screen, and participants had to press a button on the keyboard to initiate each trial. 350-850ms after participants started a trial, they were presented with a cloud of black and white dots (34 dots in total, 17 black). The positions of the dots were randomized on each trial to limit adaptation. The white fixation cross remained on the screen during the stimulus presentation, and participants were instructed to fixate at the cross throughout the entire duration of the trial. 250ms after the disappearance of the dots, the fixation cross turned green to prompt participants to reproduce the duration of the stimulus by keeping a button on the keyboard down for the entire apparent duration of the interval. Between 200-400ms after the response, participants were provided with visual feedback signifying the signed difference between the response and the actual stimulus duration. The feedback was presented for 100ms, and after the feedback disappeared, the screen remained blank for 300ms. After that, the fixation cross was presented again to signify to participants that they could start the next trial.

The duration reproduction task was divided across two sessions. The order of the conditions (short range and long range) was counterbalanced across sessions and across participants. Within each session, participants had to take a 15-minute break between the two conditions. At the start of each condition, participants completed a training block of 35 trials (7 repetitions per stimulus), in which the durations were presented in ascending order. After training, participants had to complete 5 blocks (50 trials in each block) in which the Spearman correlation between reported and actual duration was at least 0.65, and the slope was at least 0.5. If participants failed to reach the threshold at the end of a given block, they were required to complete an additional block. Participants were allowed to “repeat” a maximum of 3 blocks. If a participant failed to reach the performance criteria in four blocks in either condition and in either session, the testing session was terminated, and the participant’s data were excluded from the analysis. To summarise, depending on their performance, participants completed between 5 and 8 blocks per condition per session. In total, the data used for the analysis consisted of 100 trials per stimulus per condition.

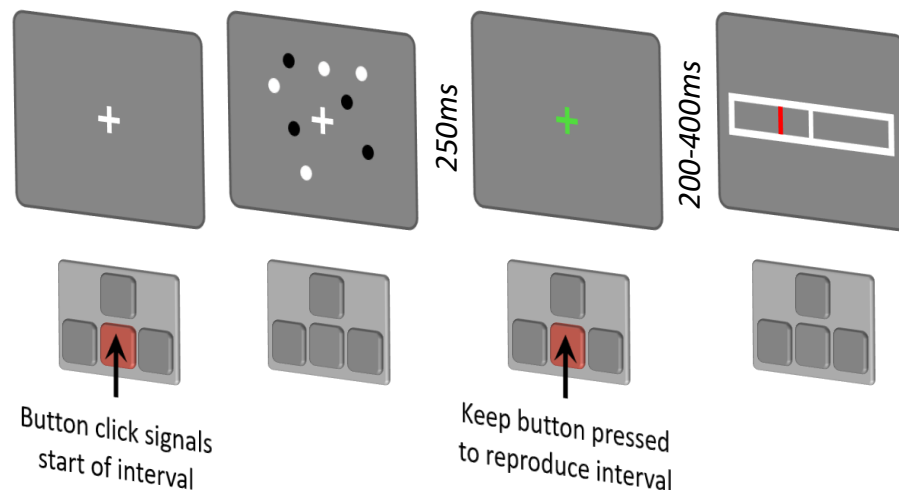


Figure 5.4. Duration reproduction task. On every trial, participants are presented with an array of black and white dots that stays on the screen for some time. 250ms after the stimulus disappears, participants are prompted to reproduce the apparent duration of the stimulus by keeping a button pressed for the same duration as the stimulus. 200-400ms after participants give their response, they are provided with feedback reflecting their accuracy. If the coloured line is to the left of the midline, then participants underestimated the presented duration, if the coloured line is to the right of the midline – participants overestimated the presented duration. If the coloured line is green, the response was within 10% of the presented duration.

Finally, trials in this task were excluded and replaced online if: (1) responses were shorter than half the duration of the shortest stimulus in the range, or longer than 1.5 times the longest duration in the range, 170-939ms in the short range and 313-1728ms in the long range; (2) participants responded too late – they did not begin their responses within 1500ms of the

presentation of the response prompt; (3) participants responded too early – they began their response before the presentation of the response cue.

2IFC task

In addition to the duration reproduction task participants also completed a 2-interval forced-choice task designed to estimate participants' sensory precision (Figure 5.5). On every trial, participants were presented with two temporal intervals – a standard and a comparison interval. The standard interval was 626ms long, and the comparison interval was, in different blocks, either shorter or longer than the standard interval. Accordingly, depending on the condition, the participants' task was to indicate which of the intervals was longer or shorter. We used the Bayesian adaptive psychometric method QUEST (Watson & Pelli, 1983) to estimate participants' 75% threshold. If on any trial the difference between the main and comparison stimuli suggested by QUEST was smaller than 0, the presented difference was drawn from a uniform distribution between 6 to 18ms.

Originally, in the preregistration, we determined that each participant would complete one block per condition (70 trials per block) and that we will estimate sensory precision by averaging the two directional thresholds. However, we decided to increase the number of blocks to two per condition per person, in order to achieve more accurate estimates of sensory precision. 2 participants have completed only one block per condition, while the rest of the participants have completed two blocks per condition.

We did not evaluate whether QUEST has converged at the end of each block but during piloting, we established that QUEST stabilized within 50-60 trials. Plots showing the convergence of the threshold for each participant are included in Appendix F (Figure F1). For the vast majority of cases, the estimate of the 75% threshold stabilized by the end of each block.

The two conditions in the 2IFC task were divided across the two experimental sessions, and they were completed after the duration reproduction task. The 2IFC condition in which participants had to indicate the longer of the two durations was completed in the session in which participants finished the duration reproduction task with the short condition. The 2IFC condition in which participants had to indicate the shorter of the two durations was completed in the session in which participants finished the duration reproduction task with the long condition.

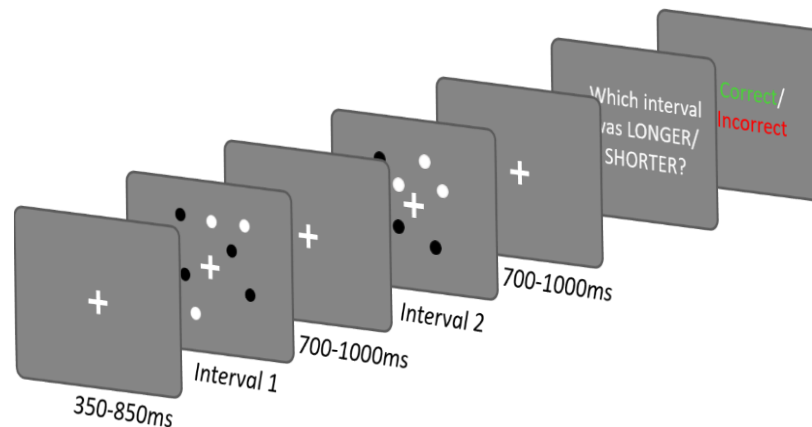


Figure 5.5. 2IFC task. On every trial, participants were presented with two intervals, and in different blocks they had to indicate which interval was shorter or longer

EEG task

In the EEG session, participants were presented with the same stimuli as in the duration reproduction task, but here they were not required to provide duration reports (Figure 5.6). On 10% of the trials (excluding trials showing the 626ms), the fixation cross changed colour from white to green, and participants were required to press a button as soon as they detected the change. The change-detection task was implemented to make sure participants were paying attention to the stimulus. Trials in which the fixation cross changed colour or participants pressed the button even if the fixation cross did not change were excluded from the analysis.

In the behavioural reproduction task, we accounted for any potential order effects by asking participants to complete two experimental sessions in which we switched the order of the conditions. Dividing the EEG task into two sessions would have caused issues for the analysis, so participants had to complete the whole task in one session. Therefore, the EEG task was divided into 4 sections, each consisting of 250 trials (5 blocks of 50 trials each). The order of the sections was either: (1) Long – Short – Short – Long, (2) Short – Long – Long – Short. Each section began with a learning block in which participants were presented with the relevant durations in ascending order so that they learn what distribution the stimuli would be drawn

from. Whenever there was a change in the condition between two sections (at the end of sections 1 and 3), participants had to take a 15-minute break.

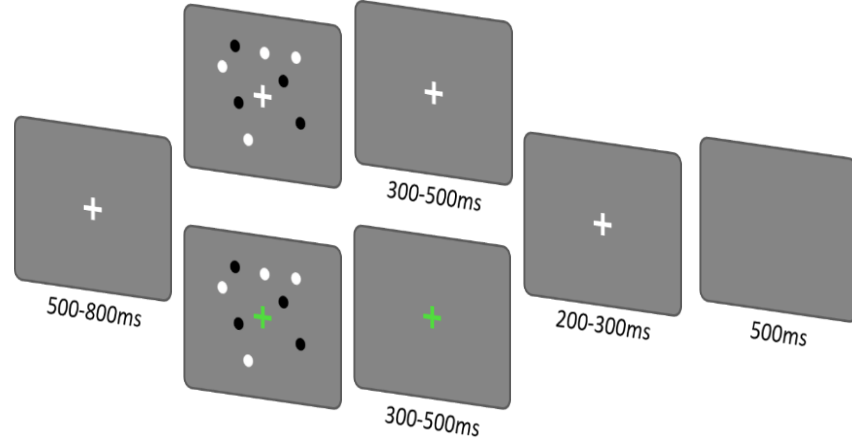


Figure 5.6. EEG task. In the EEG session, participants were presented with the same durations as in the duration-reproduction task but passively. To make sure participants were paying attention, they were required to detect changes in the colour of the fixation cross, which occurred on 10% of the trials.

Bayesian ideal-observer model

We used a modified version of the Bayesian ideal observer described by Acerbi et al. (2012) to predict participants' responses in the behavioural duration reproduction task and estimate the overall expected regression to the mean effect in the sample as well as the expected effect of sensory precision on regression to the mean. For a description of the model, please refer to Chapter 3.2 (*Model 2*, p.48).

The model has two parameters that are free to vary – sensory noise, w_s , and motor noise, w_m . To estimate the sensory noise associated with each stimulus level, we calculated Weber's fraction, WF , based on participants' performance in the 2IFC task such that $WF = threshold/626$. From there, the sensory noise associated with each stimulus level was calculated as $w_s(i) = WF * x(i)$, where x is the physical duration of stimulus i .

To fit the motor noise parameter, we used the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970). The procedure was carried out as follows. First, we determined a starting value for the motor noise parameter: 31.3ms (~5% of 626ms) and ran the model. Then we calculated the sum of squared errors between participants' responses and the responses predicted by the model for each of the two ranges (short: 340-626, long: 626-1052):

$$SS = \sum_{i=1}^5 \frac{(ResponseShort_i - PredictionShort_i)^2}{2 \times \sigma_{short}^2} + \sum_{j=1}^5 \frac{(ResponseLong_j - PredictionLong_j)^2}{2 \times \sigma_{long}^2}, \quad (24)$$

where σ_{short}^2 refers to the variance in participants responses in the short condition, and σ_{long}^2 refers to the variance in participants responses in the long condition.

From this, the probability of the data given the current motor noise parameter (or the likelihood) becomes:

$$P(\text{Response}|\text{current motor noise}) = \exp(-SS). \quad (25)$$

In the next step of the algorithm, we proposed a new value for the motor noise parameter. This value was drawn from a Gaussian distribution centred at the current motor noise value, with SD = 10. If the sampled motor noise value was smaller than 12ms or larger than 125ms, another value was sampled. We then run the model with the new proposed value and calculate the sum of squared errors between the participant's responses and the model predictions with the new motor noise value. Then we calculated the ratio between the likelihood of the data given the current motor noise value and the likelihood of the data given the new motor noise value:

$$\frac{P(\text{Response}|\text{new motor noise})}{P(\text{Response}|\text{current motor noise})} = \exp(-SS_{new} + SS_{old}). \quad (26)$$

Finally, to decide which parameter value to keep, we sampled a random number between 0 and 1. If the likelihood ratio calculated above was greater than the randomly sampled number, we replaced the original motor noise estimate with the newly proposed estimate. If the likelihood ratio was smaller than the randomly sampled number, we discarded the newly proposed motor noise estimate and kept the original estimate. The process was repeated 1000 times, and in the end, we chose the motor noise estimate that had the highest likelihood.

Behavioural statistical analyses

We used Bayesian t-tests to investigate the behavioural hypothesis of interest. In all t-tests, the null hypothesis was defined as a point distribution at 0 difference, and the alternative hypothesis was defined as a half-normal distribution with a mean of 0 and a standard deviation equal to the effect predicted by the Bayesian ideal-observer model detailed in Chapter 3. To determine the prior expected effect for the overall regression to the mean effect (responses to the 626ms will be longer in the Long condition compared to the Short condition), we computed the average model-expected regression to the mean effect across all participants. To determine the prior expected effect for the effect of sensory precision on regression to the mean (observers with greater sensory precision will exhibit weaker regression to the mean), we computed the difference in model-predicted regression to the mean between participants

whose behaviourally-measured sensory precision (as measured in the 2IFC task) was lower than the median sensory precision and participants whose behaviourally-measured sensory precision was greater than the median. Bayes factors (BF_{10}) larger than 3 were interpreted as providing evidence for the alternative hypothesis, and BF s smaller than $1/3$ were interpreted as providing evidence for the null hypothesis (Dienes, 2014).

EEG data collection and pre-processing

EEG data were recorded using a 64-channel ANT neuro amplifier at 1024Hz. EEG data analysis was conducted in the EEGLAB toolbox (Delorme & Makeig, 2004) running on Matlab2019a. The data were downsampled to 512Hz. Then the data were band-pass filtered at 0.1-45 Hz and 1-45Hz. The second filtered dataset was used to detect eye movements, and eye blinks using independent component analysis. After filtering, long breaks between blocks were deleted from the data so that artefacts due to participants movement during breaks do not affect bad channel detection. Bad channels were detected using the `clean_rawdata()` function and removed from both datasets. Next, the data were epoched from -250 to +1550ms relative to stimulus onset and manually inspected to remove bad epochs. Epoch rejection was kept to a minimum and focused on removing trials with unique artefacts and trials in which participants likely did not pay attention to the stimulus. This included epochs in which the participants blinked at the onset or offset of the stimulus, or blinked multiple times during the presentation of the stimulus. Epochs in which blinks occurred at other times were not removed. Epochs were also rejected if there was uncharacteristic muscle activity during stimulus presentation, which would signal that participants were not paying attention to the stimulus. On average across all participants and stimuli, 4.58% of epochs were rejected ($SD = 3.6$, $Max = 16.02$). Across all participants, we retained an average of 96.53 epochs ($SD = 4.83$, $Min = 79$) containing the 626ms stimulus in the short condition, and an average of 95.53 epochs ($SD = 6.43$, $Min = 82$) containing the 626ms stimulus in the long condition. Next, we ran ICA, using the Picard algorithm (Ablin, Cardoso, & Gramfort, 2018), on the dataset filtered at 1-45Hz, and projected the ICA weights to the 0.1-45Hz data set. We then used IClablel (Pion-Tonachini, Kreutz-Delgado, & Makeig, 2019) to remove eye components and components defined primarily as muscle activity (with a threshold set at 90%). Finally, the missing channels were interpolated, and the data was referenced to average.

Decoding analyses

The data was first downsampled to 128Hz. The decoding analyses were performed using the Amsterdam Decoding and Modelling toolbox version 1.07 (ADAM – Fahrenfort, Van Driel, Van

Gaal, & Olivers, 2018). An LDA classifier was trained to distinguish stimulus context (short from long) based on the EEG activity associated only with the 626ms stimulus. All EEG channels were used in the analysis. The classifier was trained and tested using 10-fold cross-validation, and classification performance was evaluated using the area under the curve (AUC). T-tests were used to assess whether classification performance was significantly better than chance (.5) at each time point in the epoch, and we corrected for multiple comparisons using cluster-based permutations.

CNV analysis

To investigate the effect of context on the brain activity associated with the overlapping stimulus (626ms) we also conducted ERP analyses using the Mass Univariate ERP Toolbox (Groppe, Urbach, & Kutas, 2011a, 2001b). The EEG activity measured by each electrode was averaged across all participants, separately for each stimulus level not used in the main analysis (340, 396, 461, 537, 729, 849, 989, and 1152ms). Electrodes were selected for analysis based on whether they displayed CNV-like activity. We defined CNV-like activity as a prolonged increase in negativity of the EEG amplitude starting after the onset of the presented interval (Figure 5.7A). Across the stimuli, we found a consistent cluster of 20 electrodes displaying CNV-like activity in parietal and occipital areas: CP6, CP4, P6, P5, P4, P3, Pz, P2, P1, P5, PO8, PO7, PO6, PO5, PO4, PO3, Poz, O2, O1, Oz (Figure 5.7B). Figure 5.7C shows the averaged ERP for each of the stimuli used to define the electrode cluster of interest.

We defined the start of the CNV as the peak with the highest positive amplitude following stimulus onset, which was not followed by another positive peak for at least 200ms. The end of the CNV was defined as the latency of the first peak with negative amplitude following the start of the CNV (see Figure 5.7A). With these guidelines, we computed the start and the end of the CNV at each electrode, and for each stimulus, averaged across participants. Next, across these latencies, we found the earliest recorded CNV start latency and the latest recorded CNV end latency. We used these values to define our CNV window in order to be as liberal as possible and to make sure we do not exclude any potentially relevant EEG activity from the analysis. The chosen CNV windows spanned from 275ms following stimulus onset to 250ms following stimulus offset.

To determine if the EEG activity associated with the 626ms stimulus presented in the long compared to short context differed within the pre-determined electrode and latency window we performed a permutation test with 2500 permutations based on paired t-tests on every time point between 275 to 876ms from stimulus onset at every electrode from the identified

cluster. Results were corrected for multiple comparisons using tmax permutations, which ensures correction for familywise error to the level of Bonferroni correction (Bonferroni test-wise $\alpha = 8 \times 10^{-6}$). This method for correction for multiple comparisons provides strong control for Type I errors and allows us to make statistical inferences regarding the latency and location of an effect. However, this method also increases the chance of Type II errors, and it disregards the fact that EEG activity extends across time and space – the activity recorded at each time point and channel is not independent of the activity recorded at neighbouring time points and channels (Bullmore et al., 1999). An alternative method for correcting for multiple comparisons is to use cluster-based permutation which takes into account the temporal and spatial clustering of EEG data. However, with this approach, no inference can be made regarding the latency and location of an effect within a cluster (Groppe, Urbach, & Kutas, 2011; Sassenhagen & Draschkow, 2019). The goal of the mass univariate analysis conducted here was to define how the effects of context and sensory precision on CNV activity extend through space and time. In addition, we wanted to minimize Type I errors, and for these reasons, the tmax permutation approach was deemed as more appropriate.

Finally, we also performed a permutation test based on independent samples t-tests to determine if there was a significant difference in context effects between the group with high and low sensory precision. The electrode cluster, latency window, number of permutations, and the method for correcting for multiple comparisons were kept the same as above.

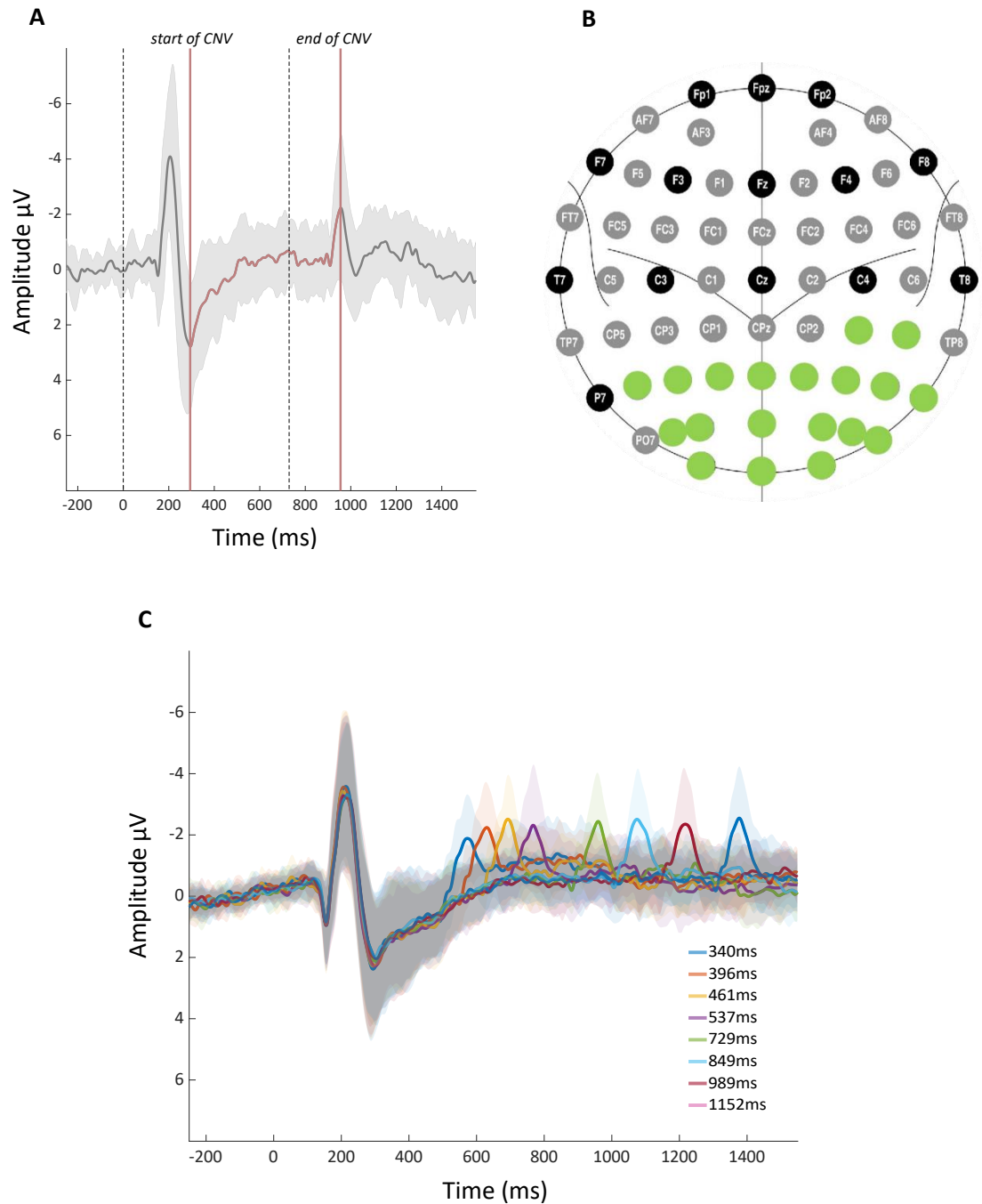


Figure 5.7. CNV latency and electrode cluster. (A) Example of CNV-like activity. Averaged activity for stimulus duration 729ms at POz. The black dotted lines show stimulus onset and offset, and the red lines show the CNV window. The shaded areas show mean amplitude \pm 1SD. **(B)** Cluster of electrodes selected for CNV analysis. Electrodes marked in green displayed CNV-like activity in response to the 340, 396, 461, 537, 729, 849, 989, and 1152ms stimuli. The EEG data associated with the 626ms stimuli was not included for the selection of electrodes. **(C)** ERP activity averaged across all selected electrodes in the cluster for each stimulus duration. The shaded areas show mean amplitude \pm 1SD.

Participants

The final sample for this experiment consisted of 20 participants (Mean age = 25.14, SD age = 3.69). One participant was excluded from the EEG analysis due to an experimenter error, the participant did not complete a sufficient number of blocks in one of the conditions. To arrive at our final sample, we tested a total of 48 participants. However, 26 participants were excluded because they did not fulfil the performance criteria (see *Procedure: Duration reproduction task*, p.84-85), while 2 additional participants dropped out.

The required sample size ($N=20$) was determined by running the behavioural statistical analysis detailed above on duration reproduction data simulated with the Bayesian ideal observer model. More specifically, we determined the sample size expected to provide a sensitive result, $BF_{10}>3$, for the analysis of the effect of sensory precision on regression to the mean. To simulate the needed data, we had to define the sensory and motor noise parameters for each “participant”. Sensory noise was varied between 10-20% of the stimulus duration (Rammsayer & Pichelmann, 2018; Rammsayer, 2014; Lapid et al., 2008), while motor noise was kept constant at 64ms. To mimic the structure of the experiment, in each simulation, we generated 500 trials per condition. The simulations revealed that we could expect to find evidence of a difference in regression to the mean between participants with low and high sensory precision ($BF_{10}>3$) with a sample of 20 participants (10 participants per group). Briefly, the simulations revealed that we can expect an average regression to the mean effect of 98.57ms ($SE=1.80$) and that this effect will be 59.57ms ($SE=7.15$) stronger in observers with low sensory precision (also see Baykova & Roseboom, 2019).

5.2.2 Behavioural results

Confirmatory analyses

Overall regression to the mean effect

The first aim of the behavioural experiment was to investigate if, overall, the regression to the mean effect predicted by a Bayesian ideal observer model can provide a good explanation for participants’ behavioural responses. Across all participants, the model predicted an overall regression to the mean effect of 128.43ms (Figure 5.8A). The regression to the mean effect in the behavioural data was 89.78ms ($SE=7.90$, Figure 5.9A), which, even though numerically smaller than the effect predicted by the model, resulted in extreme evidence in favour of the alternative hypothesis ($BF_{10} = 1.14 \times 10^{27}$). Therefore, these results provide compelling evidence that regression to the mean effects in duration perception can be described by Bayesian

observer models. Finally, in the behavioural data, regression to the mean ranged between -16.08 to 137.71ms (IQR = 31.39). A histogram showing the distribution of regression to the mean across participants is included in Appendix E (Figure E1).

Effect of sensory precision on regression to the mean

The main aim of our experiment was to examine if sensory precision has an effect on regression towards the mean. Namely, our hypothesis was, that greater sensory precision will be associated with weaker regression towards the mean. We used a median split to separate participants into low and high sensory precision groups based on their performance on the 2IFC task. Based on this division, the Bayesian ideal observer model predicted that the regression to the mean effect in the low sensory precision group would be 69.82ms greater than the regression to the mean effect in the high sensory precision group (Figure 5.8B). In the behavioural data, however, the regression to the mean effect in the low sensory precision was 5.82ms (SE = 4.73) *higher* than the regression to the mean effect in the high sensory precision group, and these results provided very strong evidence in favour of the null hypothesis ($BF_{10}=0.0193$, Figure 5.9B).

Exploratory analyses

After carrying out the confirmatory analyses specified in our pre-registration, we decided to conduct additional exploratory analyses to further assess some issues identified after collecting the data. On average, participants' performance in the 2IFC task was in line with previous findings (e.g. Rammsayer, 2014) and there was variability in the estimated thresholds between participants (Mean Weber Fraction = 0.19, SD = 0.08, IQR = 0.12). A histogram showing the distribution of thresholds across participants is included in Appendix F (Figure F2). However, the estimated Weber fractions ranged from 0.09 to 0.38, with 5 participants having an average Weber fraction greater than 0.25. We found these results suspicious because if participants' thresholds were so large, they should not have been able to complete the duration reproduction task. To deal with this issue, we decided to exclude blocks in which the threshold was greater than 25% of the main stimulus when estimating the sensory precision of each participant. A histogram showing the distribution of corrected thresholds across participants is included in Appendix F (Figure F3). The corrected thresholds, represented as Weber fractions, had a mean of 0.14 (SD = 0.05), and ranged between 0.03 to 0.24 (IQR = 0.04). Another issue we identified was that one participant did not show regression to the mean – on average, their responses to the 626ms stimulus were 16ms shorter in the long compared to the short condition. The model we used in this analysis cannot produce repulsive

effects, and it would, therefore, not provide a good account for their data. This is further emphasized by the fact that the best fitting value for the motor noise parameter was very close to the upper boundary of the range in which we allowed motor noise to vary (124.7, with 125 being the maximum allowed value). This occurs because in the model when sensory noise is fixed, regression to the mean is diminished by increasing motor noise. Without the upper boundary set for the motor noise parameter, the best fit that this model can provide for data with repulsive effects will be achieved with values approaching infinity. Therefore, we decided to run the same analyses as above, but this time excluding the participant that does not exhibit regression to the mean.

Employing these new exclusion criteria did not provide us with vastly different results. The behavioural data again supported the overall regression to the mean effect predicted by the model: average model predicted bias = 95.81ms, average behavioural bias = 95.35ms ($SE=5.41$), $BF_{10} = 1.58 \times 10^{66}$. Nonetheless, we still found evidence of no difference in regression to the mean between the low and high sensory precision group. The model predicted that the bias exhibited by the group with low sensory precision would be 65.28ms higher than the bias exhibited by the group with high sensory precision, while the difference in the behavioural data was 5.26ms ($SE = 3.99$, $BF_{10} = 0.26$).

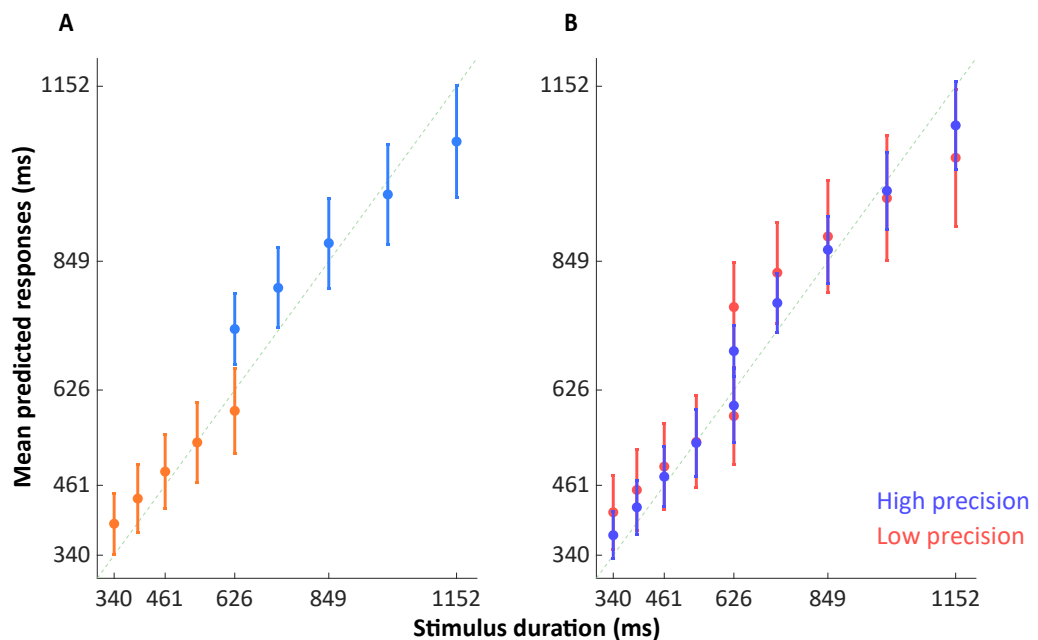


Figure 5.8. Predictions of the Bayesian ideal observer model. (A) Overall regression to the mean effect predicted by the model. **(B)** Effect of sensory precision on regression to the mean predicted by the model. The green dotted line in both figures represents veridical responses. The error bars represent the average standard deviation of the predicted response probability distribution across participants associated with each duration.

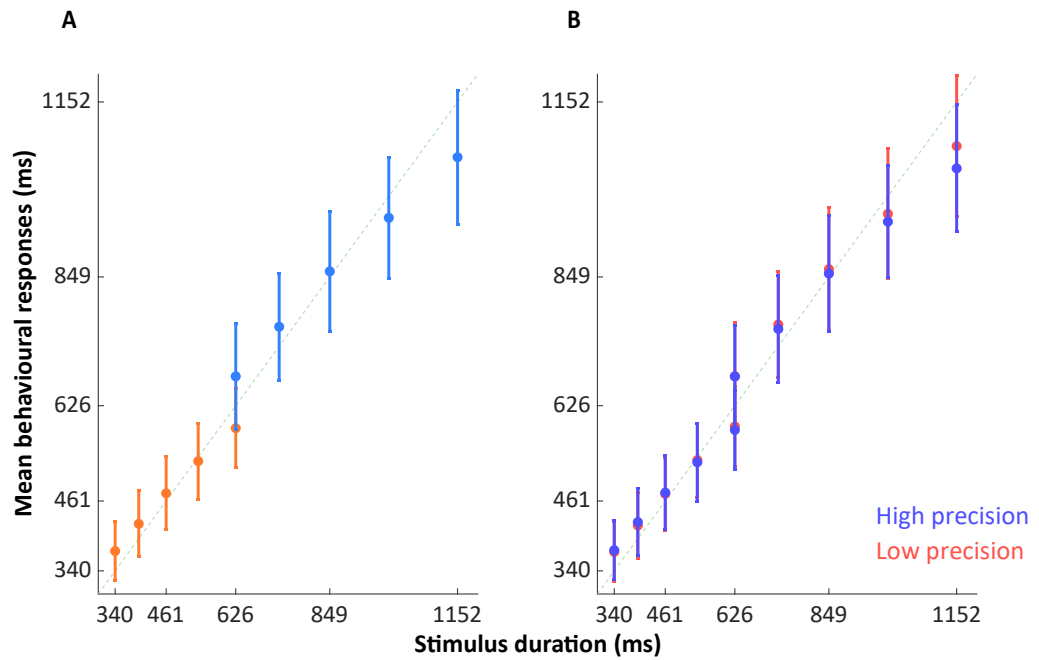


Figure 5.9. Effects in the behavioural data. (A) Overall regression to the mean effect in the behavioural data. **(B)** Effect of sensory precision on regression to the mean in the behavioural data. The green dotted line represented in both figures represents veridical responses. The error bars represent the average standard deviation of participants' responses to each duration.

5.2.3 EEG results

Context decoding

The first aim of the EEG experiment was to determine if the stimulus context influences the neural activity associated with the same physical stimulus. We found that the classifier was able to discriminate context, short stimulus range versus long stimulus range, based only on the EEG activity associated with the 626ms stimulus (Figure 5.10). The results reveal two clusters in which classification performance was significantly better than chance: 187.7-225.6ms, peak AUC = 0.551, $p < .05$; and 242.2-281.3ms, peak AUC = 0.550, $p < .05$. As the gap between the two clusters is only 2 samples, 15.6ms, it is likely that data is highly variable, and this is in fact, one cluster. These results suggest that the neural activity associated with a given stimulus does not reflect solely the physical characteristics of that stimulus but also the context in which this stimulus is presented, or in Bayesian terminology, the prior. As the latencies of the identified clusters (or cluster) show, the EEG activity that is most meaningful in discriminating context across participants takes place very shortly after stimulus onset, at which point participants still do not know what stimulus they are being shown – the shortest possible duration presented in the experiment was 340 ms. As a result, we suggest this activity reflects information predominantly regarding the prior.

Sensory precision and context decoding

Above we presented evidence that stimulus context is reflected in the EEG signal associated with the same physical stimulus. Another question we wanted to investigate is whether sensory precision has an influence on this effect. We predicted that low sensory precision would amplify the effect of context on the neural processing of the 626ms. In other words, we hypothesized that a context classification based on the EEG data associated with the 626ms would be better in participants with low sensory precision.

First, we treated sensory precision as a categorical variable, and we split the participants into a high and a low sensory precision group using a median split. As an estimate of decoding performance, we extracted the peak AUC for each participant across the entire epoch. Contrary to our prediction, we found no difference in decoding performance between the participants with high sensory precision (mean peak AUC = 0.665, SE = 0.012) and the participants with low sensory precision (mean peak AUC = 0.656, SE = 0.012).

In a second analysis, we treated sensory precision as a continuous variable and checked if there is a correlation between sensory precision and decoding performance. Here we found a negative correlation between sensory *noise* and AUC, $r = -0.18$. So, as sensory *noise* increases (i.e., sensory *precision* decreases), decoding performance decreases. These results go against our prediction that as sensory precision decreases, regression to the mean would increase, resulting in a greater influence of stimulus context on the neural processing of the 626ms stimulus.

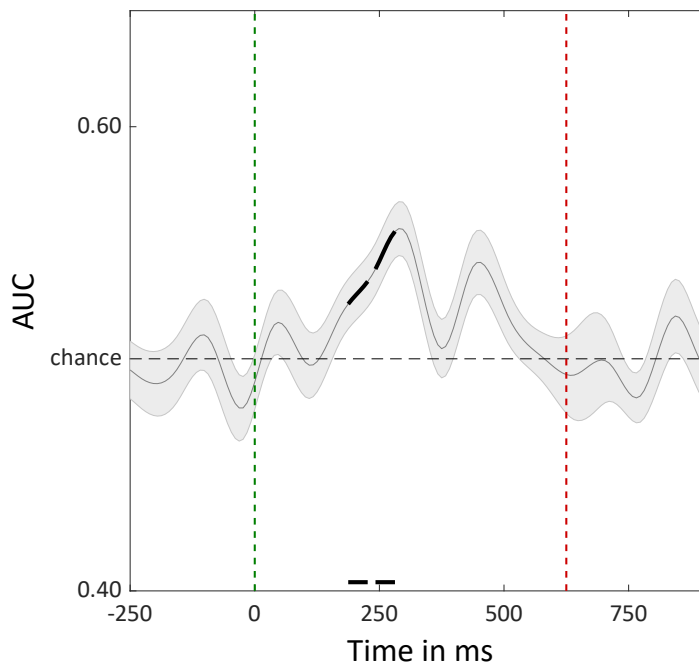


Figure 5.10. Classification performance (AUC) across time, averaged over participants. *The green dotted line represents stimulus onset; the red dotted line represents stimulus offset. The highlighted areas from 187.7-225.6ms and from 242.2-281.3ms represent the clusters in which performance is significantly better than chance. The shaded areas represent the standard error of the mean calculated across participants.*

Regression to the mean and context decoding

Finally, we wanted to investigate if regression to the mean is correlated with context classification performance. We predicted that as regression to the mean increases, the neural representation of the 626ms stimulus would become more and more influenced by context, which would result in better classification performance. However, we found no correlation between regression to the mean and peak AUC ($r = -0.04$).

CNV results

The mass univariate analysis showed some significant differences in ERP amplitude between the overlapping stimulus, 626ms, when presented in the two different contexts. Specifically, the ERP amplitude at P1 from 730.47 to 740.23ms (significant corrected p-values between 0.030 and 0.045) relative to stimulus onset was significantly more positive in the long compared to short context condition (Figure 5.11). However, there were no significant differences in the effect of context (626Long – 626Short) on brain activity differences between the group with high and low sensory precision (all corrected p-values > 0.869).

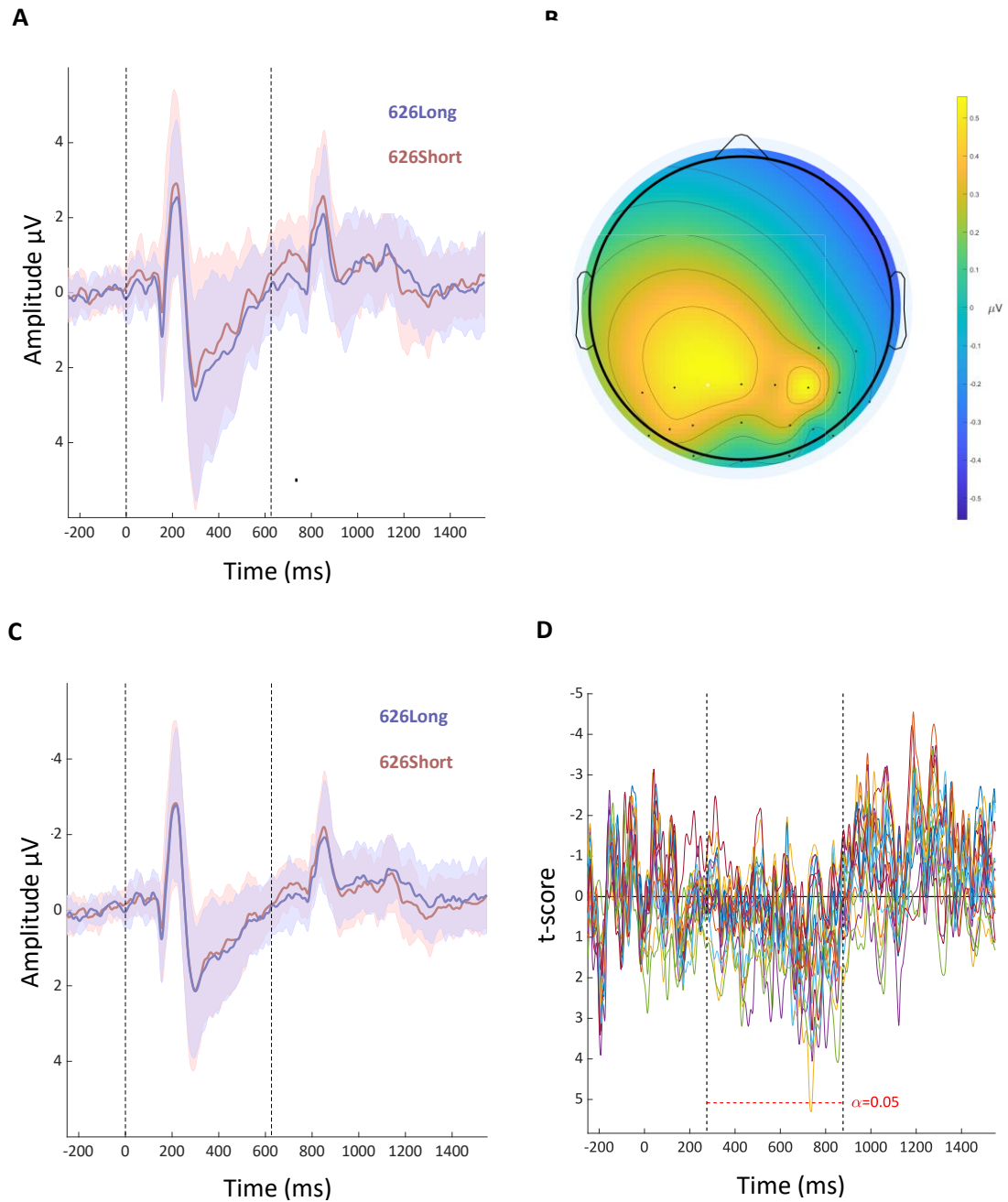


Figure 5.11. ERP results (A) ERP at P1 in response to the 626ms stimulus at each context. The dotted lines show stimulus onset and stimulus offset. There is a significant difference in ERP amplitude from 730.47 to 740.23ms. The shaded areas show mean amplitude \pm 1SD. (B) Topography of the difference between 626Long and 626Short at 730.5ms. (C) ERP in response to the 626ms stimulus at each context averaged across all electrodes in the cluster. The dotted lines show stimulus onset and stimulus offset. The shaded areas show mean amplitude \pm 1SD. (D) T-score as a function of time. Each coloured line is a different electrode from the pre-specified electrode cluster. The dotted black lines show stimulus onset and stimulus offset. The dotted red lines represented the critical t-value (after correction for multiple comparisons) – t-values exceeding the critical value are significant.

5.2.4 Discussion

In this experiment, we successfully replicated the general regression to the mean effect reported in other duration estimation studies and found evidence of a potential EEG-based correlate of this effect. Focusing on the behavioural results, we found that the same physical stimulus was overestimated when it was shorter than the mean of the distribution, and underestimated when it was longer than the mean of the distribution, which is in line with previous studies in this area (for example Lejeune & Wearden, 2009; Jazayeri & Shadlen, 2010, 2015). The regression to the mean effect in the behavioural data was consistent with the regression to the mean predicted by the Bayesian ideal observer model, providing evidence that these perceptual biases could arise from a Bayesian-like process in the brain.

Furthermore, our EEG experiment provided evidence for a potential neural correlate of the regression to the mean effect. We found that a classifier was able to distinguish above-chance whether the same physical stimulus came from a distribution of short or a distribution of long durations based on the EEG data associated with that stimulus alone. Observers were not required to report the durations of the stimuli, suggesting that this effect could be predominantly perceptual. Similar results were recently reported by Damsma et al. (2020), who used an auditory interval reproduction task, providing further evidence for the generalizability of this effect across tasks and modalities. Together, these results provide strong evidence that the brain activity elicited during the processing of a given stimulus reflects the overall context in which this stimulus is presented. It is, therefore, possible to use classification performance as a proxy measure to assess the magnitude of perceptual biases on a “neural” level similarly to how regression to the mean can be used to assess the magnitude in perceptual biases in behaviour. Finally, we found a limited effect of context on the amplitude of the CNV associated with the same physical duration, but our results contradict previous findings. While Damsma et al. (2020) found that the amplitude of the CNV was more negative when the same stimulus was presented amongst longer compared to shorter durations, we found that the amplitude of CNV was more negative when the same stimulus was presented amongst shorter durations. The differences in these results could be associated with the differences in the implemented tasks – participants in our study observed the stimuli passively, while participants in the study by Damsma et al. (2020) had to reproduce the presented durations. Previous research has provided evidence that the CNV is composed of an early and late component, with the late component showing a greater increase in negativity in tasks involving or allowing greater motor preparation (Gaillard & Näätänen, 1973; Loveless & Sanford, 1974; MacKay & Bonnet, 1990; Vidal, Bonnet, & Macar, 1995). However, to date, no

studies have investigated to what extent the response mode used in duration estimation tasks affects the characteristics of the CNV. If there are substantial differences in CNV amplitude between time perception tasks involving differences in motor preparation, for example, direct estimation compared to duration reproduction, the role of the CNV in time perception would have to be revisited, and further research would be needed to disambiguate the sensory, decision, and motor processes that contribute to the CNV.

The main aim of the present experiment was to investigate if sensory precision affects the magnitude of behavioural and neuroimaging perceptual biases towards the prior.

Behaviourally, a Bayesian ideal observer model predicted that observers with low sensory precision would exhibit stronger regression to the mean than observers with high sensory precision. Based on this, we also predicted that in observers with low sensory precision, the brain activity associated with a given stimulus will be more influenced by the overall stimulus context compared to observers with high sensory precision. In other words, we predicted context classification (short vs long range) based on the EEG data associated with the same physical stimulus would be more accurate in observers with low sensory precision.

Nonetheless, in contrast to previous studies (Cicchini et al. 2012; Aagten-Murphy et al. 2014), the behavioural results provided strong evidence that sensory precision had no influence on regression towards the mean and the EEG results found no evidence of a difference in classification performance or difference in CNV amplitude between the two groups.

It is important to note that the change-detection task we used to ensure that participants were paying attention to the stimuli presented in the EEG session does not ensure that participants were attending to the stimuli's duration. It is possible that participants could have completed the change-detection task without attending to the duration of the stimuli. Nonetheless, when defining the electrode cluster of interest for the mass univariate analysis, we observed that EEG activity differed between the durations presented in the study (**Figure 5.7C**). While this is no indication of whether participants paid attention to the intervals, it provides us with some assurance that the brain engaged in some stimulus-related activity. If we had found no differences in brain activity between stimuli of different durations, that would have been a strong indicator that participants were not engaged in the task. In addition, previous studies have found that changes in CNV activity can be elicited during passive observation when stimulus-related responses are not required (Donchin, Gerbrandt, Leifer Tucker, 1972; Ruchkin, Sutton, Mahaffey, & Glasser, 1986), while changes in MMN can be elicited even when participants are not attending to the stimulus characteristic manipulated in the experiment (e.g. Näätänen, Paavilainen, Rinne, & Alho, 2007). However, the role of attention in estimating

durations remains an open question (for example, consider the differences between retrospective and prospective timing tasks, e.g. Zakay & Block, 1996), and it is very likely that in the current study there were individual differences in how participants were engaging with the stimuli during the EEG session which could have increased the heterogeneity of the results. Therefore, to ensure stricter control, future studies would benefit from employing a task that specifically directs participants attention towards the duration of the stimuli, such as asking participants to report the duration of the intervals presented on a selection of the non-critical trials.

One possible issue in our experiment is that the motor noise associated with reproducing temporal intervals could be masking the relationship between sensory precision and regression towards the mean. The relationship between sensory noise and regression to the mean proposed by the model is straightforward. If a participant has low sensory noise, their internal estimates of the presented durations will be close to veridical. If a participant has high sensory noise, their internal estimates of the presented durations will be biased towards the mean. The motor noise parameter in the model, however, affects responses in a different way. If a participant has low motor noise, they will be able to produce their desired response accurately. If a participant has high motor noise, they will not be able to produce their responses accurately, and responses will be shifted away from the optimal command. In fact, high motor noise dampens the regression to the mean effect. If we take two participants with the same sensory noise, but different motor noise, the participant with high motor noise will, overall, overestimate the presented duration – their point of indifference will be higher than the mean of the distribution (Figure 5.12). As a result of this shift, the participant with high motor noise will underestimate the longest durations in a range less than the participant with low motor noise. This effect, however, is not uniform across the entire range of presented durations, and switches for very short stimuli – a participant with high motor noise will overestimate short durations *less* than a participant with low motor noise. When we combine these two effects – less overestimation of short intervals and less underestimation of long intervals – participants with low motor noise will exhibit weaker regression towards the mean when compared to a participant who has the same internal estimate but lower motor noise. Furthermore, in extreme cases, the model can predict similar regression towards the mean for participants who have low sensory and low motor noise, and participants who have high sensory and high motor noise. This suggests it is possible that the motor noise associated with reproducing durations has masked the relationship between sensory precision and regression towards the mean in the behavioural data. This could similarly explain why we did not find a

correlation between the regression to the mean effect measured in the behavioural sessions, and the performance of the classifier trained to distinguish context – the EEG task involved passive observation, and the measured effect was not influenced by additional noise, while the behavioural effect was affected by motor noise.

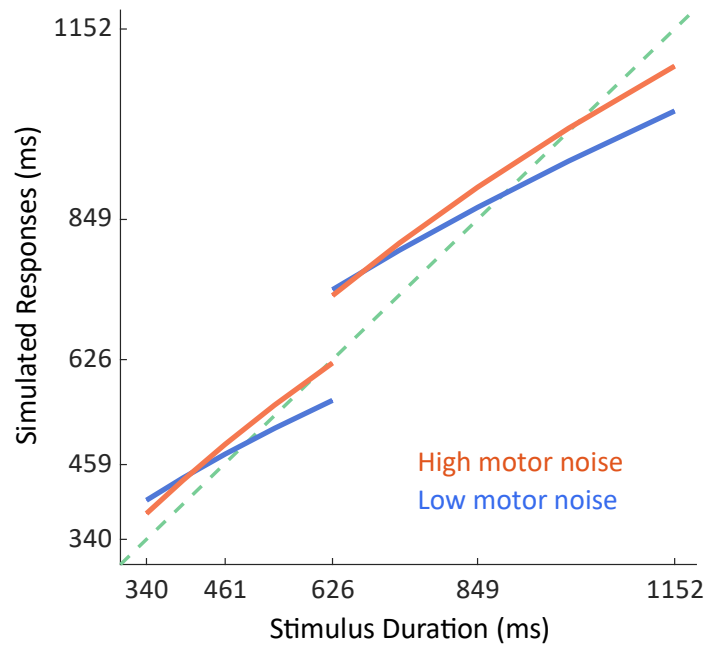


Figure 5.12. Effect of motor noise. *Model-generated responses of two simulated participants with the same sensory noise but different motor noise. The participant with high motor noise (red) tends to underestimate the longest durations in each range less than the participants with low motor noise (blue). In addition, the participant with high motor noise tends to overestimate the shortest durations in each range less than the participant with low motor noise. As a result, the difference in responses to the 626ms stimulus between the two ranges is smaller for the participants with high motor noise.*

While motor noise could have masked the effect of sensory precision on regression to the mean, it is not clear why we did not find an effect of sensory precision on classification performance. It is possible that our measure of sensory precision was unreliable – some participants showed high variability in discrimination thresholds between different blocks of the 2IFC task. Therefore, we will have to replicate this experiment, taking caution to obtain more reliable estimates of sensory precision before we can draw any strong conclusions.

5.3 Experiment 2

To address the issues identified in Experiment 1, we ran an online behavioural experiment in which we investigated the relationship between regression to the mean and sensory precision in the absence of motor noise confounds by using a direct estimation rather than a

reproduction task. In the direct estimation task, participants were asked to report the durations of the stimuli on a visual analogue scale, and this allowed us to more closely examine how sensory precision affects perceptual biases. This experiment was originally planned to be a full pre-registered EEG replication of Experiment 1 in which we were going to resolve the other identified issues, but we had to adapt our plans due to the COVID-19 pandemic.

5.3.1 Method

Pre-registration and data availability

The hypotheses, design, experimental procedure, and analysis plan for the behavioural experiment are pre-registered at the Open Science Framework (<https://osf.io/bmsc7/>). The data is available here: <https://osf.io/bmsc7/files/> (this includes data from the simulations to determine sample size, and the behavioural data).

Design

In this experiment, we investigated whether sensory precision affects the magnitude of perceptual biases towards the prior in a behavioural duration estimation task. As in Experiment 1, sensory precision was quantified as participants' average 75% threshold in a 2IFC duration discrimination task. Regression to the mean in the duration estimation task was defined as the difference in participants' responses to the same physical stimulus, 626ms, presented in a range of short (340-626ms) or long (626-1152ms) durations. Our prediction was that as the 75% threshold increases (i.e., as sensory precision decreases) regression to the mean should increase.

Apparatus

The behavioural experiment was run on PsychoPy (Peirce et al., 2019). Participants completed the experiment remotely on their own computers, so there were variations in size, resolution, and refresh rate. Participants were asked to input the size of the screen, and to sit ~57 cm away from the screen. The main stimulus was a stationary cloud of 17 black and 17 white dots presented on a grey background. The size and duration of the stimuli were defined based on the characteristics of each participant's monitor to ensure that the consistency of the stimuli was stable across participants. Each dot had a size of 0.5dva, and the overall size of the entire cloud was 12dva. The position of each dot was randomized on every trial to mitigate visual adaptation effects. Participants provided responses using standard computer keyboards and mice.

Procedure

Participants completed 2 sessions in total. Each session comprised of a duration estimation task and a duration discrimination task and took ~2 hours to complete. Instructions for all tasks can be found in Appendix G.

Duration estimation task

The estimation task followed the same structure as the duration reproduction task in Experiment 1. The main difference was that instead of reproducing the duration of the presented stimulus, participants had to report the stimulus duration on a visual analogue scale. In each experimental session, participants completed two blocks from each condition – a training block and a testing block. The order in which the conditions were presented was counterbalanced across sessions and participants, and there was a 15-minute mandatory break between the two conditions. In the training blocks, participants were presented with 5 repetitions of 8 durations that were not part of the testing blocks (short condition training durations: 271, 316, 369, 429, 500, 582, 678, and 790ms; long condition training durations: 500, 582, 678, 790, 920, 1072, 1249, and 1454 ms). Testing blocks consisted of 250 trials or 50 trials per duration, with 5 durations in each condition (short condition test durations: 340, 396, 461, 537 and 626ms; long condition test durations: 626, 729, 849, 989, and 1152ms).

At the beginning of each trial, a fixation cross was presented along with a prompt asking participants to start the trial by pressing a button on the keyboard. The fixation cross remained on the screen throughout the entire duration of the trial. 600-1000ms after participants started the trial, the stimulus was presented, and after the stimulus disappeared, the fixation cross remained on the screen for a further 250ms. After that, the visual analogue scale appeared, and participants had to provide a response by dragging a computer mouse along the scale. The scale ranged from half the duration of the shortest stimulus in the range to 1.5 times the longest duration in the range – 170-939ms in the short range and 313-1728ms in the long range. In the testing blocks, participants were allowed to opt out instead of providing a response. Participants were allowed to skip a maximum of 10 trials. If participants provided a response, the feedback was presented for 1000ms. The feedback display was the same as in Experiment 1. Finally, 200-500ms after the feedback disappeared, a prompt was presented, asking participants to start the next trial. If participants failed to start a trial within 15s of the presentation of the prompt or failed to provide a response within 5s of the presentation of the response screen, the trial was marked as missed. If participants missed more than 10 trials, the experiment ended. Finally, if there was more than a 10ms difference between the desired and

achieved duration of the stimulus, the trial was excluded. If more than 20 errors in achieved stimulus duration were detected during a block, the experiment was terminated. Trials with errors in achieved duration, missed, and skipped trials were reshuffled back into the remaining stimulus sequence.

At the end of every testing block, we computed the Spearman correlation and slope between the presented and reported duration of the stimuli. If participants did not reach a correlation of 0.65 and a slope of 0.5, the experiment was terminated.

Duration discrimination task

On every trial of the duration discrimination task, participants were presented with two stimuli and had to report which stimulus was shorter or longer in different blocks. The task followed a similar structure to the 2IFC task used in Experiment 1, but here, the difference in the stimuli presented on every trial was defined using the method of constant stimuli rather than QUEST. The pre- and post-stimulus intervals varied between 500-800ms and the interval between the two stimuli varied between 600-1000ms. Participants had 15s to provide a response after which they were presented with feedback regarding the accuracy of their responses for 1s. After the disappearance of the feedback, participants were prompted to start the next trial.

The task began with a training block consisting of 16 trials in which the two presented intervals were 426 and 826ms long. In the testing blocks, the main stimulus had a duration of 626ms, and the comparison stimulus was either shorter or longer than the main stimulus. In each condition, there were 6 comparison stimuli which differed by 5-30% (in steps of 5%) from the main stimulus. Each comparison stimulus was repeated 15 times per block, and participants completed between 2 and 4 blocks per condition depending on their performance. At the end of each block, a cumulative Gaussian psychometric function was fitted to participants data. The values estimated during the course of the experiment were used only as an indication of participants' performance and were not used in the analysis. If the 75% threshold estimated by the model exceeded 187ms, 30% of the main stimulus' duration, the block was discarded, and participants were asked to complete an additional block up to a maximum of 4 blocks. If the threshold exceeded 187ms in all blocks, we used the last two blocks in the analysis. Skipped, missed trials, and trials in which there was more than a 10ms difference between desired and actual duration were repeated. If participants missed more than 10 trials, or there were more than 20 trials with stimulus duration errors in a block, the experiment ended.

After participants completed the experiment, we used the Palamedes toolbox (Prins & Kingdom, 2009) in Matlab to fit a cumulative Gaussian psychometric function to participants'

data and assess its goodness-of-fit (PAL_PFML_Goodness_OffFit). Our plan, detailed in the pre-registration, was to fit a psychometric function to each block within a condition separately, exclude blocks in which pDev was less than 0.05 and exclude participants if both blocks in either condition do not meet the goodness-of-fit criterion. For the purposes of this work only, we fitted the psychometric function to the combined data of all blocks together, and participants were excluded if the goodness-of-fit for both conditions was too low. If only one condition had to be excluded due to poor fit, the participant's estimate of sensory precision was based only on the 75% threshold associated with the condition showing a good model fit.

Analysis and Bayesian ideal-observer model

We calculated Pearson's r correlation between participants' measure of sensory precision and regression to the mean and conducted a Bayesian t-test to test whether the correlation in the behavioural data supports the correlation predicted by a Bayesian-ideal observer model. The null and alternative hypotheses for the Bayesian t-test were defined in the same manner as in Experiment 1.

We used a Bayesian ideal-observer model to simulate participants' responses to the duration estimation task and estimate the expected correlation between sensory precision and regression towards the mean. Here we used a Bayesian ideal-observer model with an updating prior because during initial simulations we found that the model with a fixed prior (*Chapter3: Model 2*, p.48) produced data in which the correlation between sensory precision and regression towards the mean was unrealistically high (Pearson's r between 0.9-1). In contrast to the model with a fixed prior, a model with an updating prior (*Model 3*, p.50) produced correlations more in line with our expectations given the results of Experiment 1 (results from 100 simulations with 30 participants each: Mean Pearson's $r = 0.43$, SD = 0.16).

The sensory noise for each stimulus was estimated based on participants' average 75% threshold in the 2IFC task following the procedure detailed in Experiment 1 (p.88). Report noise was fitted using the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970), following a similar procedure as the parameter fitting of motor noise was in Experiment 1 (p.88). The starting value for report noise was 5ms, and report noise was allowed to vary between 1-25ms. Because we use a model with an updating prior, the model will not predict the same responses to each stimulus as in the model with fixed prior used in the previous experiment. Therefore, the sum of squared errors between participants' responses and the responses predicted by the model were calculated on a trial-by-trial basis. The proposed value

of the report noise parameter was sampled from a Gaussian distribution centred at the current report noise value, with a standard deviation of 5ms.

Participants

The minimal sample size for this experiment was determined by simulating behavioural responses to the duration estimation task using the same Bayesian ideal-observer model that we subsequently used to analyse the collected data. We ran simulated experiments with sample sizes varying between 30 and 60 participants to check what sample size would be needed to find conclusive evidence for the alternative hypothesis ($BF_{10} > 3$; 100 simulations for each sample size). Each simulated participant had a Weber fraction between 5-25% and a report noise of 12.5ms. The prior expected correlation was set to 0.2, as this was our minimal correlation of interest. The simulations revealed that we could expect to find evidence in favour of the alternative hypothesis ($BF_{10} > 3$) with a sample of 30 participants. We, therefore, aimed to recruit a minimum of 30 participants before running the analysis described above.

To date, 93 participants have signed up to complete the study. Of them, 40 participants completed the study and sent their data back to us, and of them, 27 ($Mean_{Age} = 31.85$ years, $SD_{Age} = 10.51$) fulfilled the inclusion criteria specified in the *Procedure* section above (p.106-108). This sample includes 2 participants who do not have a full dataset for the duration estimation task and 2 participants who do not have a full dataset for the duration discrimination task. Although this sample is below our target, running the planned analysis will not have any detrimental effects on the validity of the drawn statistical inference, so here we present preliminary findings based on the available data.

5.3.2 Results

Confirmatory analysis

This experiment aimed to investigate if, as predicted by Bayesian frameworks, regression to the mean would be correlated with sensory precision. Our hypothesis was that as sensory precision decreases, the magnitude of perceptual biases towards the prior should increase. Therefore, we predicted that participants' 75% thresholds in the 2IFC task would be positively correlated with the effect of context (long vs short) on participants' responses in the duration estimation task, and we used a Bayesian ideal-observer model to derive a specific numerical prediction of the expected effect.

As in Experiment 1, we found that in the duration estimation task, participants' responses to the same physical stimulus, 626ms, were on average 103.29ms ($SE = 11.87$) longer when the

stimulus was presented in the range of long, compared to the range of short durations. Estimates of regression to the mean ranged from -84.30 to 231.20ms (IQR = 71.09). A histogram showing the distribution of regression to the mean across participants is included in Appendix E (Figure E2). Results from the 2IFC task also revealed some variability in performance across participants, with estimates of the Weber fraction ranging from 0.07 to 0.22 between participants (Mean Weber Fraction = 0.13, SD = 0.03, IQR = 0.04). A histogram showing the distribution of thresholds across participants is included in Appendix H.

Finally, we found a positive correlation between this context effect and participants' 2IFC thresholds, $r = 0.14$ (SE = 0.2; red line in Figure 5.13). However, this relationship was weaker than the correlation predicted by the Bayesian ideal-observer model, $r = 0.63$, and consequently, the data provided inconclusive evidence in favour of the null hypothesis, $BF_{10} = 0.80$.

Exploratory analysis

In this experiment, we had very strict exclusion criteria to make sure that participants' estimates of sensory precision are not confounded by fatigue and lapses of attention. Nonetheless, exploring the results presented in **Figure 5.13**, some researchers would argue that two of the data points are potential outliers and could be biasing the results. According to the interquartile method of outlier detection, one participant had an outlying regression to the mean value (-84.3ms), and another participant had an outlying 2IFC threshold (Weber Fraction = 0.22). Therefore, we conducted the same analysis as above, but this time excluding these values, and we found that the correlation between sensory precision and regression to the mean increased to 0.27 (SE = 0.21, grey line in **Figure 5.13**), which provided inconclusive evidence in favour of the alternative hypothesis, $BF_{10} = 2.03$.

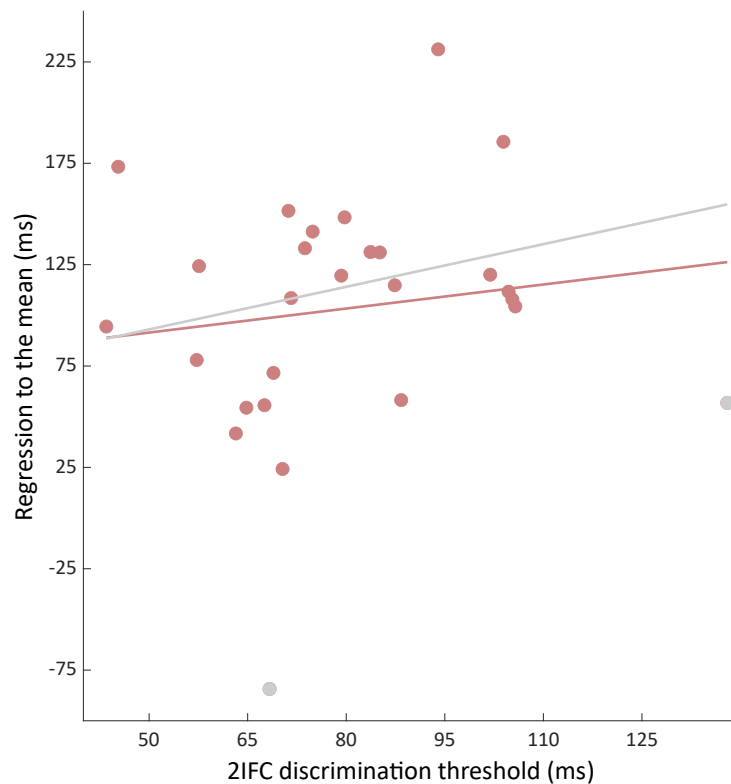


Figure 5.13. Correlation between sensory precision and regression to the mean. *Each dot represents one participant. In grey is the data of the two participants defined as outliers in the exploratory analysis. The red line shows the best-fit line associated with the confirmatory analysis; the grey line shows the best-fit line after removing the potential outliers.*

5.3.3 Discussion

Here we present a study investigating the correlation between sensory precision and regression to the mean in a direct duration estimation task. In Experiment 1, we found that, when treated as a categorical variable, sensory precision had no effect on perceptual biases towards the prior in a duration reproduction task. However, it was possible that these results were confounded by the effect of motor noise on reproduced durations, and a lack of strict data exclusion criteria when estimating sensory precision. While these issues were taken into account in the design of Experiment 2, we did not find support for the relationship predicted by the Bayesian ideal observer model. The results of our pre-registered analysis, though inconclusive, suggest that the correlation in the behavioural data is weaker than the correlation predicted by the model and point towards the null hypothesis. Applying additional exclusion criteria to remove outliers in the data changes the direction of the evidence in the behavioural data in favour of the alternative hypothesis, but the evidence is again inconclusive. The results of both analyses suggest that we need to continue data collection, and as these

exclusion criteria were not part of our original pre-registration, the results of the exploratory analysis will need to be further replicated in an independent sample.

Bayesian brain theories suggest that the variance of the sensory likelihood should affect the magnitude of perceptual biases – the smaller the variance of the sensory likelihood (so the greater the sensory precision), the weaker the bias towards the prior. However, we did not find evidence in favour of this prediction across two experiments. In Experiment 1, we found strong evidence suggesting that sensory precision does not affect regression to the mean and our confirmatory analysis in Experiment 2 found inconclusive evidence in favour of the null hypothesis. One interpretation of these results is that indeed sensory precision does not influence regression towards the mean. However, an alternative possibility is that our measure of sensory precision in the duration discrimination task did not reflect participants' sensory precision in the duration estimation task.

Here we use the term sensory precision as synonymous with sensory sensitivity. In 2IFC tasks, sensory sensitivity is reflected in the threshold of the psychometric curve, also known as the just-noticeable difference – the minimal change in stimulus magnitude required to elicit the perception of a new stimulus (Kingdom & Prins, 2010). Participants with lower thresholds will differentiate stimuli that participants with higher thresholds will not be able to. Discrimination thresholds have been used as a measure of sensory precision in previous studies. For example, it was found that discrimination thresholds were related to the width of the likelihood function in a Bayesian model of speed perception (Stocker & Simoncelli, 2006b). In addition, studies on multisensory cue combination have used the thresholds from unimodal discrimination tasks to determine the width of the sensory likelihood associated with each unimodal signal (e.g. Ernst & Banks, 2002). However, one limitation of using the threshold as a measure of sensory precision is that the threshold is affected by both the observers' sensitivity to changes in the stimulus and the observers' internal variability (Kingdom & Prins, 2010; Vilidaite, Yu, & Baker, 2017). The presentation of a stimulus elicits a response in the brain, and an observer is going to discriminate between two stimuli when there is sufficient difference in the brain responses they elicit. The same stimuli, however, will not always give rise to the same neuronal responses – brain activity fluctuates with time, causing a stimulus difference that was discriminable on one trial to not be discriminable on the next (Milne, 2011). As a result, an observer with low sensitivity and low internal variability might have the same discrimination threshold as an observer with higher sensory sensitivity but also higher internal variability.

Given these issues with interpreting the threshold and the fact that we did not find evidence of a relationship between the discrimination threshold and performance in the estimation task, we investigated whether there is a relationship between regression to the mean and a different 2IFC performance metric – the slope of the psychometric curve fitted to the data. In forced-choice tasks, the slope of the psychometric curve is inversely proportional to the variability of participants' discrimination performance, with steeper slopes indicating a greater increase of correct responses as the difference between the stimuli increases (Kingdom & Prins, 2010). It is important to note, however, that the slope is influenced by both sensory and decision noise – while the threshold conflates sensory sensitivity and internal variability, the slope conflates sensory sensitivity and response variability. Nonetheless, if the precision of responses in the discrimination task is related to bias in the estimation task, we would expect that increase in the slope of the psychometric function fitted to the 2IFC data will be related to a decrease in regression to the mean in the estimation task. In contrast to this prediction, we found that the correlation between regression to the mean and the slope was $r = -0.08$ (after removing 1 outlier, $r = -0.04$). These results suggest that both the threshold and slope of the psychometric function are not ideal measures of sensory precision, and future studies would benefit from employing additional tasks to estimate internal and response noise in order to provide an unconfounded measure of sensory sensitivity.

Another possibility for the findings discussed above is that the process underlying duration discrimination differs from the process underlying duration estimation, and consequently, the lack of relationship between regression to the mean and 2IFC performance metrics stems from insufficient transfer across tasks. Anecdotally, some participants reported that throughout the duration discrimination task, their strategy changed from comparing the presented durations to identifying the main stimulus – they were able to select their response after seeing the first stimulus. Therefore, the task changes from a duration discrimination task into a memory task. An alternative measure of sensory precision used in the literature is the coefficient of variation which is calculated directly from participants' responses in magnitude estimation tasks as the ratio between the standard deviation of participants responses to a given stimulus and the magnitude of that stimulus (e.g., Gibbon, 1977; Wearden 1991). This measure of sensory precision is taken directly from the duration estimation task itself, so there would be no problems with task transfer (although an issue with this approach is that regression to the mean and the coefficient of variation are jointly estimated). Furthermore, we found that participants' coefficient of variation (averaged across all stimuli in the two conditions of the duration estimation task) was not correlated with their 75% thresholds in the 2IFC tasks,

$r=0.05$ (after removing 1 outlier the correlation was $r = -0.12$), or the slope of the psychometric curve in the 2IFC task ($r = 0.02$). Therefore, we investigated whether our original measure of regression to the mean (the effect of context on responses to the 626ms stimulus), was correlated with participants' average coefficient of variation. Across all participants, the correlation was -0.26 , and after removing 1 outlier, the correlation was 0.05 . These additional analyses show that in our sample, there was almost no relationship between two measures of sensory precision which in the literature are used to determine the same psychophysical construct.

As we found that neither the 75% threshold in the 2IFC task nor the coefficient of variation in the duration estimation task correlated with regression to the mean in the direction predicted by Bayesian brain frameworks, it is possible that the context effect in the duration estimation task does not arise as a result of Bayesian inference. An alternative account that could give rise to a similar effect is discussed in Chapter 8 (*General Discussion: Wider implications*, p.145). We also decided to explore whether either measure of sensory precision would correlate with a different measure of regression to the mean which is often used in the literature – the regression index ($1 - \text{slope of the best fit line}$; Cicchini et al., 2012; Aagten-Murphy et al., 2014; Barne et al., 2018). The regression index of each participant was based on the average slope across the four experimental blocks in the duration estimation task. We found a correlation of 0.46 between participants' regression index and coefficient of variation, and a correlation of 0.19 (or 0.3 after removing 1 outlier) between participants' regression index and 75% 2IFC threshold. The discrepancy between all these results suggests that further research is necessary to determine the construct validity of different measures of sensory precision and perceptual biases before investigating the relationship between them. Establishing a measure's validity on face value alone is not enough as our knowledge of the potential underlying processes contributing to a specific effect is limited.

Chapter 6. Attractive biases in duration estimation outside of awareness

The experiments presented in the previous chapters provide evidence that the perceptual system tracks the statistical properties – range, central tendency, and variance – of distributions of temporal intervals. However, it is still unknown whether conscious awareness is necessary to estimate the duration of stimuli or to assimilate statistical information across time. Here I present an experiment in which we investigated whether the duration of stimuli masked using continuous flash suppression give rise to local (serial dependence) and global (regression to the mean) attractive biases in duration estimation. Participants ($N=5$) were asked to estimate the duration of visual stimuli ranging between 150 to 750ms. In two different conditions, we used continuous flash suppression to mask either the two shortest or two longest stimuli in the range thus creating a dissociation between the mean of all presented stimuli and the mean of the visible stimuli in the two conditions. If masked stimuli did not contribute to regression to the mean, then responses in each condition would be biased towards the mean of just the stimuli visible in that condition. First, we found significant differences in participants' responses between the two conditions, suggesting that the masked stimuli did not contribute to the estimation of stimulus summary statistics, but participants' responses were biased against rather than towards the mean of the visible stimuli. We also found no significant difference in participants' responses following the different masked stimuli suggesting the durations presented outside of awareness do not give rise to serial dependence effects. However, these results are undermined by our inability to show strong evidence of serial dependence following the visible durations, which suggests that our study was underpowered. Overall, these findings provide evidence that the estimation of durations requires conscious awareness, and suggest that stimuli presented outside of awareness likely do not contribute to Bayesian inference. However, further research will be necessary to investigate what factors gave rise to the negative bias we observed on a global level.

6.1 Introduction

Studies have found evidence that sensory information of varying complexity can be processed outside of awareness, but it is still not clear where the limits of unconscious perception lie. Adaptation effects have been found to survive masking in low-level features such as orientation (e.g. He & Macleod, 2001) and motion (e.g. Roumani & Moutoussis, 2020), and even higher-order features such as numerosity (Liu, Zhang, Zhao, Liu, & Li, 2013). The research

on unconscious perception of more complex stimuli, such as faces, has however provided contradictory findings. Some studies report activation of face-processing brain areas in response to invisible faces (Moutoussis & Zeki, 2002; Morris, Pelphrey, & McCarthy, 2007), while others report no sensory adaptation to facial expression (Yang et al., 2010) and identity (Moradi et al., 2005).

No studies to date have investigated if duration perception requires awareness or not, and different theories provide conflicting predictions. It has been proposed that magnitude judgements, such as size, numerosity, and duration, share a common mechanism (Meck & Church, 1983; Walsh, 2003), which would suggest that if numerosity can be perceived outside of awareness, so should duration. However, studies have challenged the proposition of (completely) overlapping representation, by showing, for example, only unidirectional interference between numerosity and duration in perceptual tasks (Droit-Volet, Clément, & Fayol, 2003; Dormal, Seron, & Pesenti, 2006). Theories on how the brain estimates durations also provide conflicting predictions regarding the role of awareness in time perception. On the one hand, in models based around scalar timing theory, conscious awareness is not necessary for the internal clock to operate. Even updated versions of the theory which suggest a role of attention (e.g., Zakay & Block, 1996; Lejeune, 1998) into the internal clock do not necessitate awareness as attention and awareness can be dissociated (e.g., Hsieh, Colas, & Kanwisher, 2011; Lami, Alon, Carmel, & Shalev, 2014). In contrast, models which suggest that time perception depends on perceptual content (e.g., Roseboom et al., 2019) would predict that conscious awareness of the presented stimulus is necessary for its duration to be estimated.

The experiment presented here investigated if conscious awareness is necessary for the duration estimation of visual stimuli. Furthermore, we also investigated whether durations presented outside of awareness contribute to the formation of prior expectations locally, in serial dependence effects, and globally, in regression to the mean effects. To date, very few studies have investigated if stimuli presented outside of awareness contribute to the estimation of summary statistics (Suárez-Pinilla, 2018; Kim et al., 2020). Notably, Suárez-Pinilla (2018) found that in an orientation estimation task, masked stimuli did not give rise to serial dependence effects but contributed to the estimation of the mean orientation in the stimulus set. Here, following a very similar design, we presented participants with a distribution of durations and used continuous flash suppression (CFS) to mask either the shortest or longest stimuli. First, if masked stimuli generate serial dependence effects, we can expect that stimuli following the short masked durations will be underestimated, while stimuli following the long masked durations will be overestimated, and this difference would be similar to the effect of

the equivalent non-masked durations. Regarding the predicted regression to the mean effect, if the masked stimuli contribute to the computation of mean duration, duration estimates would be biased towards the mean of all stimuli in the distribution irrespective of which stimuli are masked. Conversely, if the masked stimuli do not contribute to the computation of mean duration, we predict that estimates would be biased towards the mean of only the visible stimuli.

6.2 Methods

6.2.1 Data availability

The behavioural data is available at: <https://osf.io/j9fqg/>.

6.2.2 Design

Across two conditions participants were presented with 9 durations between 150 and 750ms in steps of ~75ms, with a mean of 450ms (150, 230, 300, 380, 450, 530, 600, 680, 750ms). In one condition, the two shortest stimuli, 150 and 230ms, were masked, shifting the mean of the visible stimuli up to 527ms. In a second condition, the two longest stimuli, 680 and 750ms, were masked which shifted the mean of the visible stimuli down to 377ms.

First, to investigate if participants could estimate the duration of stimuli without awareness, we focused on the effect of masking on the absolute error between participants' estimates and the actual duration of the four stimuli (150, 230, 680 and 750ms). If awareness is necessary to estimate duration, then the absolute error between responses and stimuli will be greater when the stimuli are masked than when they are visible.

To test whether visual stimuli presented outside of awareness contribute to regression to the mean effects we focused on the effect that masking those stimuli had on participants' direct duration estimates of the stimuli visible in both conditions (300-600ms). If durations presented outside of awareness contribute to the computation of mean stimulus magnitude, duration reports in both conditions would be biased towards the mean of all presented stimuli – 450ms. If durations presented outside of awareness are not considered in such computations, reports would be biased towards the mean of the visible stimuli in each condition, 527ms in condition 1 and 377ms in condition 2. This would also be the case for masked stimuli, so we can also predict that responses to the masked stimuli in condition 1 (150 and 230ms) will be longer than responses to the masked stimuli in condition 2 (680 and 750ms).

Finally, to investigate whether awareness is necessary for serial dependence effects in duration estimation, we tested the effect of the duration presented on the previous trial (trial $n-1$) on the standardized error between the participants' response and the presented duration on the current trial (trial n). To calculate the standardized error between the participant's response and the duration presented on every trial, we first calculated the normalized error, NE, as:

$$NE_i = \frac{\text{Reported duration}_i - \text{Presented duration}_i}{\text{Presented duration}_i} \quad (27)$$

where i denotes trial number. From there, the standardized error, SE, on trial i was calculated as:

$$SE_{i,k} = \frac{NE_{i,k} - \text{Mean}(NE \text{ on all trials showing duration } k)}{SD(NE \text{ on all trials showing duration } k)} \quad (28)$$

where k denotes the duration presented on trial i .

A negative standardized error would mean that the participant underestimated the presented duration, while a positive standardized error would mean that the participant overestimated the presented duration. If the duration presented on the previous trial affects responses on the current trial, short durations presented on trial $n-1$ would lead to negative standardized errors on trial n . Conversely, long durations presented on trial $n-1$ would lead to positive standardized errors in trial n .

6.2.3 Stimuli and apparatus

The experiment was run on Matlab2018b using Psychtoolbox-3.0.14 (Brainard, 1997, Pelli, 1997; Kleiner, Brainard, Pelli 2007). The stimuli were presented on a 22-in Iiyama Master Pro 513 CRT monitor with 100Hz refresh rate and 800x600 pixel resolution. Participants viewed the screen through a stereoscope, which separated the visual field of the left and right eye. Throughout the experiment, participants were asked to rest their head on a chinrest which was positioned such that the participants' eyes would be aligned with the centre of the screen.

The main stimulus, which was presented to the participants' non-dominant eye, was a black-and-white Gabor with size of 5 dva, spatial frequency of 0.45 cycles per degree, 15° orientation, 0.25° phase and a Gaussian envelope with sigma equal to 10 pixels. The Gabor appeared and disappeared gradually using a temporal Gaussian envelope with a standard deviation equal to 0.25 of the stimulus duration. The maximum contrast that the Gabor reached during the trial was determined based on participants' performance in the 2IFC task.

Throughout the task, the maximum contrast decreased if participants reported a conscious experience of the masked Gabor during the CFS task (more details on this below).

On some trials, a circular black-and-white mask was presented to the dominant eye of the participant. The mask had a size of 10dva and was divided into equally-sized squares (0.5x0.5 dva). Each square was assigned a random luminance level between 0% and 100%. Every 100ms, the luminance of each square in the mask was updated.

A white square frame was presented around the location of the stimulus and mask to aid the binocular alignment of the visual input presented to the left and right eye.

6.2.4 Procedure

The experiment consisted of 5 sessions. In session 1, participants completed a contrast calibration task and training in duration estimation. In sessions 2-5, participants completed a shortened version of the duration estimation training, the main CFS task, and a shortened version of the contrast calibration task. Session 1 took ~1 hour to complete and sessions 2-5 took ~2.5 hours each. Instructions for all tasks can be found in Appendix I.

Continuous flash suppression set up

Before each task, participants completed a CFS set up which aimed at determining the position on the screen where the stimulus and mask should be presented. Participants positioned their chin on a chin rest. They were then presented with two square frames to each eye. In the middle of one of the squares was a vertical line, in the middle of the other square - a horizontal line. Participants' task was to align the two images so that the frames overlap and the lines in the middle of the frames form a cross. The coordinates of the aligned frames were extracted as the coordinates where the stimuli will be shown. Once the set-up was complete, participants were advised to remain still. Participants were allowed to take breaks at the end of each block, and the setup procedure was repeated at the start of the next block.

Contrast calibration task

The contrast calibration task was a 2-interval forced-choice (2IFC) task designed to determine at what contrast participants would not be able to identify the presence of the grating behind the mask. On every trial, participants were presented with two intervals one after the other. The mask was presented to the participant's dominant eye in both intervals. In one of the intervals, a grating was presented to the participants' non-dominant eye. Participants' task was to identify which interval contained the stimulus. Both intervals were 1000ms long. The mask

was presented to the dominant eye for the entire duration of the interval. The stimulus was presented to the non-dominant eye in the middle of the interval. We tested participants' contrast sensitivity to the shortest and longest stimuli in the range, 150 and 750ms, at 5 contrast levels - 12.5%, 25%, 50%, 75%, 100%. The initial contrast calibration completed in session 1 consisted of 400 trials, 40 trials for each duration and contrast level combination. For each of the two durations, a cumulative Gaussian psychometric curve was fitted to participants data, with the threshold and slope free to vary. During pilot testing, we noticed that at contrast levels greater than 36%, the grating was easily visible through the mask. Therefore, if for either duration in the contrast calibration task, the contrast value at which participants correctly identified the interval containing the grating with 55% accuracy was greater than 36%, the initial contrast value was set to 36%. If it was lower than 36%, the initial contrast value was set to that value. The 2IFC task completed at the end of sessions 2-5 consisted of 200 trials, 20 trials for each duration and contrast combination. If the contrast level at which participants were 55% accurate at identifying the interval containing the grating had decreased, the starting contrast for the grating in the next session was updated.

The lowest possible contrast at which the stimulus in the main task was allowed to be presented was 10%. With the gradual onset and offset, if the peak stimulus contrast was below 10%, much of the stimulus was very difficult to see (even without masking), affecting duration estimates. If during the main experimental task participants reported conscious experience of the masked stimulus, the maximum stimulus contrast was decreased following a 10-step logarithmic scale between the highest and lowest allowed contrast. If participants reported conscious experience of the masked Gabor at 10% contrast, the experiment was terminated.

Duration estimation training

In the duration estimation task, participants were presented with a black and white grating and asked to report how long the grating remained on the screen for. During training, all stimuli were unmasked. Participants were presented with 21 durations from 76 to 862ms in steps of ~30-40ms (80, 110, 150, 190, 230, 260, 300, 340, 380, 410, 450, 490, 530, 560, 600, 640, 680, 710, 750, 790, 860). Each trial started with a 200ms pre-stimulus interval. Then the stimulus was presented centred in the middle of a 1000ms stimulus presentation period, and in the end, the trial finished with a 200ms post-stimulus interval. Participants were then asked to report the duration of the stimulus by moving a computer mouse to the left or the right – moving the mouse to the right increased the duration while moving the mouse to the left decreased the duration. The initial duration presented on the response screen on each trial

was randomized between 0 and 1000ms, and the longest allowed response was 1000ms. At the end of the trial, participants received feedback containing the duration of the stimulus and the duration of their response. The duration estimation training in session 1 consisted of 210 trials, 10 repetitions per stimulus. The duration estimation training in sessions 2-5 consisted of 63 to 210 trials depending on performance. The training was terminated when the Pearson correlation between the presented durations and participants' responses was at least 0.85.

CFS duration estimation task

The main experimental task was completed in sessions 2-5 and consisted of a duration estimation task in which on some trials, the stimulus was masked using CFS (Figure 6.1). The duration of the grating presented to the non-dominant eye of the participants ranged between 150 to 750ms in steps of 75ms (9 durations in total). Each block consisted of 63 trials, 7 repetitions per duration, and participants completed 7 blocks per session. Participants completed 4 sessions of the main CSF task. In two of the sessions, the two shortest stimuli were masked (150 and 230ms), and in the other two sessions, the two longest stimuli were masked (680 and 750ms). Each trial began with a pre-stimulus interval, 150-350ms, during which a square frame was presented to both eyes. Next, the stimulus was presented centred within an 800ms stimulus presentation period, and the trial ended with a 150-350ms post-stimulus interval during which only the frame was presented. If the stimulus was masked, a high-contrast dynamic mask was presented throughout the entire duration of the trial. After the end of the post-stimulus interval, participants were asked to estimate the duration of the Gabor irrespective of whether they had a conscious experience of the stimulus or not. Duration reports were given the same way as in the duration estimation training. No feedback was provided after the response. Finally, we used the Perceptual Awareness Scale (PAS, Sandberg & Overgaard, 2015) to determine if participants had a conscious experience of the stimulus. If participants reported a PAS scale of 3 ("almost clear experience") or 4 ("clear experience"), the contrast of the grating was decreased following the procedure described above.

6.2.5 Participants

Five participants ($\text{Mean}_{\text{age}} = 26.40$, $\text{SD}_{\text{age}} = 0.56$) completed the experiment.

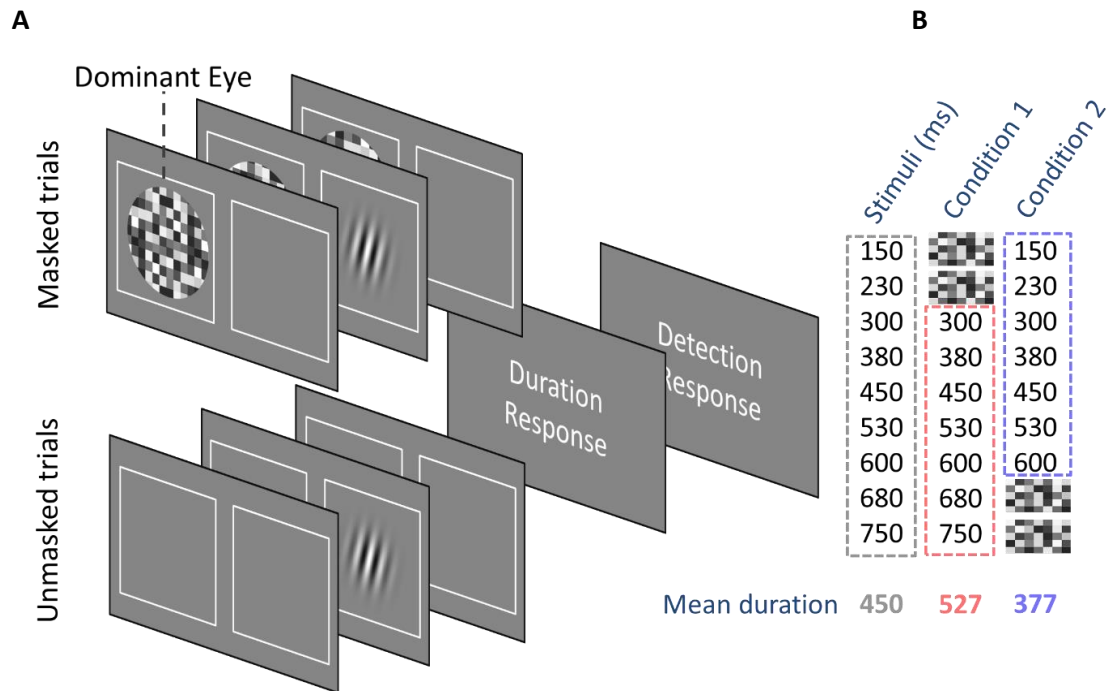


Figure 6.1. Main experimental task. (A) Trial structure. On every trial, a grating of variable duration was presented to the participants' non-dominant eye, and on some trials a dynamic high-contrast mask was presented to their non-dominant eye. Participants were instructed to report the duration and visibility of the grating irrespective of whether they were aware of it or not. **(B) Stimuli and conditions.** The durations of the gratings ranged between 150-750ms. In Condition 1, the two shortest stimuli were masked (150 and 230ms), while in Condition 2, the two longest stimuli were masked (680 and 750ms). This shifts the mean of all visible stimuli from 450ms to 527ms in Condition 1 and to 377ms in Condition 2.

6.3 Results

6.3.1 Does duration estimation require conscious awareness?

First, to investigate if participants could estimate the durations of stimuli presented outside of awareness we carried out a 2X4 repeated-measures analysis of variance (ANOVA) on the absolute error between responses and stimuli with visibility (2 levels: masked and unmasked) and duration (4 levels: 150, 230, 680, and 750ms) as independent variables. We found a significant main effect of masking, suggesting that participants' responses were more accurate when stimuli were visible than when stimuli were masked, $F(1, 4) = 56.253$, $p = 0.0017$. We also found a significant main effect of duration, $F(3,12) = 18.230$, $p = 9.154e-05$, as well as a significant masking X duration interaction, $F(3,12) = 15.432$, $p = 2.025e-04$, which indicated that the increase in duration had a significantly stronger effect on responses when stimuli were visible compared to masked.

Additional planned pairwise t-tests were then carried out to test the effect of masking on each of the four durations. We found that, on average, absolute errors were significantly higher when the stimuli were masked rather than visible for each of the four duration (**Figure 6.2**): 150ms – $t(4) = 7.140$, $p = 0.003$; 230ms – $t(4) = 4.937$, $p = 0.007$, 675ms – $t(4) = 7.299$, $p = 0.003$; 750ms – $t(4) = 6.910$, $p = 0.003$. All p-values are FDR-corrected.

6.3.2 Do durations outside of awareness contribute to regression towards the mean effects?

Next, we tested the prediction that participants responses to the stimuli overlapping between the two conditions (300-600ms) would be longer when the shortest stimuli were masked (150 and 230ms; condition 1), than when the longest stimuli were masked (680 and 750ms; condition 2). To examine if masked stimuli contributed to regression to the mean, we ran a 2x5 repeated measures ANOVA on the duration estimates of the 5 overlapping stimuli (300-600ms) across the 2 conditions (shortest stimuli masked or longest stimuli masked). We found a main effect of stimulus duration on reports such that estimates increased with the duration of the presented stimuli, $F(4,16) = 145.892$, $p = 2.260e-12$. We also found a significant effect of condition, $F(1,4) = 17.692$, $p = 0.0136$, however the results went in the opposite direction of our hypothesis - participants responses to the overlapping stimuli were longer in condition 2, Mean response = 480.996 (SD = 106.195), than in condition 1, Mean response 415.913ms (SD = 100.410). There was no significant interaction between condition and stimulus duration, $F(4,16) = 0.660$, $p = 0.629$.

We also carried out additional planned pairwise t-tests examining the effect of condition on each of the 5 durations (Figure 6.2). The difference between conditions was significant in responses to the 300, 450 and 600ms stimuli (300ms: $t(4) = 4.238$, $p = 0.045$, 95% CI [20.187, 96.903], 450ms: $t(4) = 4.174$, $p = 0.045$, 95% CI [25.923, 128.942], 600ms: $t(4) = 3.785$, $p = 0.045$, 95% CI [20.250, 131.750]). The difference between conditions was not significant for the 375ms ($t(4) = 2.964$, $p = 0.072$, 95% CI [3.689, 113.185]) and the 525ms stimuli ($t(4) = 2.750$, $p = 0.072$, 95% CI [-0.542, 110.562]). All p-values are FDR-corrected.

Finally, we also predicted that responses to the masked stimuli in condition 1 would be longer than responses to the masked stimuli in condition 2. However, responses to the masked stimuli in both conditions were closer to the mean of all durations presented in the experiment rather than the mean of only the visible durations in each condition, Mean_{Condition1} = 418.664ms (SD = 48.322), Mean_{Condition2} = 433.301 (SD = 21.648), and we found no significant difference between duration reports, $t(4) = -0.508$, $p = 0.638$, 95% CI [-94.643, 63.370].

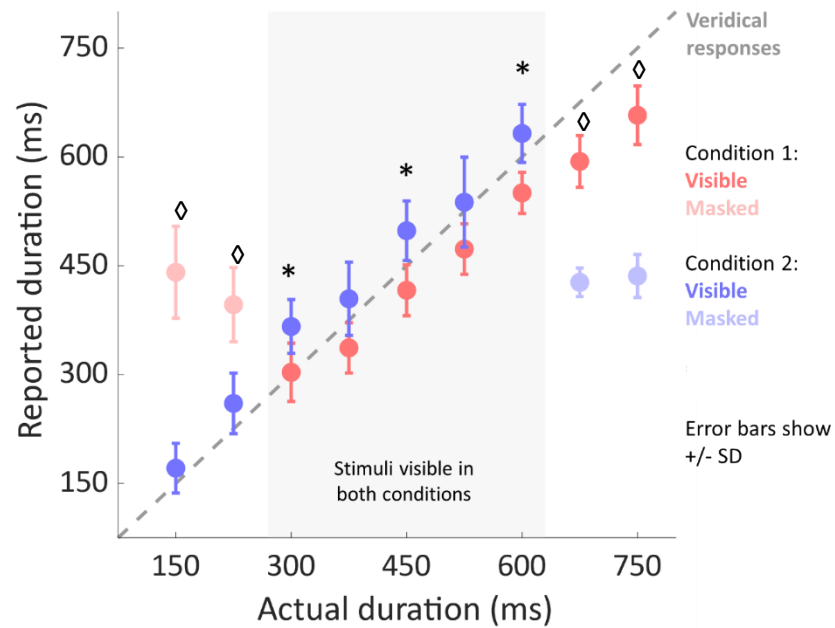


Figure 6.2. Mean responses to the stimuli in each condition. First, participants were significantly more accurate in estimating the shortest and longest stimuli (150, 230, 680 and 750ms) when they were visible compared to masked. Second, responses to the overlapping stimuli were shorter in condition 1, where the shortest stimuli were masked, than in condition 2, where the longest stimuli were masked. Focusing specifically on the stimuli visible in both conditions, the difference in responses was significant for the 300, 450, and 600ms.

◇ denotes $p < 0.05$ (FDR-corrected) for a comparison of absolute errors between stimuli.

* denotes $p < 0.05$ (FDR-corrected) for a comparison of duration estimates between stimuli.

6.3.3 Do durations presented with and without awareness give rise to serial dependence effects?

To test whether the duration of visible stimuli created serial dependence effects we fitted a linear regression between the standardized error between participants' responses and the presented duration on trial n , and the duration of the stimulus presented on trial $n-1$. Overall, participants underestimated the current stimulus when the previous trial presented a short stimulus (evidenced by negative standardized error) and overestimated the current stimulus when the previous trial presented a long stimulus (evidenced by positive standardized error). We found that in condition 1 (visible stimuli: 300-750ms) the duration presented on trial $n-1$ significantly predicted response bias on the current trial, $b = 4.332e-4$ ($SE = 9.788e-5$), $t(5) = 4.427$, $p = 0.007$ and the model provided significantly better fit to the data than the null, $F(1,5) = 19.596$, $p = 0.007$, $R^2 = 0.797$. However, in condition 2 (visible stimuli: 150-600ms) there was no significant serial dependence effect, $b = 4.92e-4$ ($SE = 2.032e-4$), $t(5) = 2.419$, $p = 0.060$, and

we failed to reject the null model, $F(1,5) = 5.851$, $p = 0.060$, $R^2 = 0.539$. The results are presented in Figure 6.3.

Finally, to investigate if masked stimuli elicited serial dependence effects, we ran a one-way repeated-measures ANOVA on the standardized error between stimulus and response on trial n with the masked duration presented on trial $n-1$ (4 levels: 150, 230, 680, and 750ms) as an independent variable. On average, three of the four masked durations lead to a slight underestimation of the following stimulus, but the analysis revealed no significant differences in responses between the 4 masked stimuli, $F(3,12) = 0.89$, $p = 0.473$.

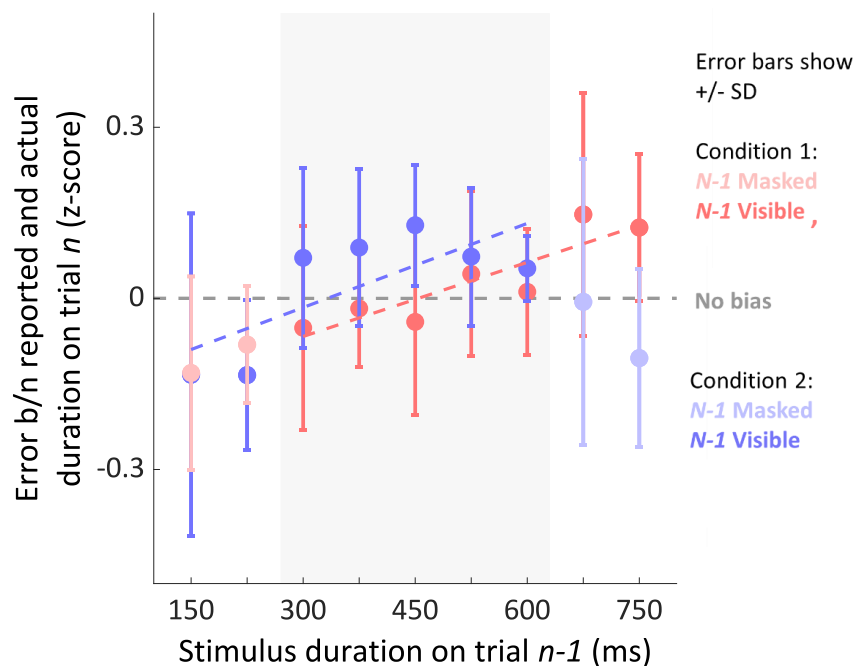


Figure 6.3. Serial dependence effects. The dotted coloured lines represent the best-fit lines for the regression between visible durations presented on trial $n-1$ and the response bias on trial n . We found a significant effect of visible stimuli on subsequent reports only in condition 1. We also found no effect of masked stimuli on subsequent responses.

6.4 Discussion

The aim of the present experiment was to investigate whether estimates of stimulus duration require conscious awareness and whether stimuli presented outside of awareness contribute to regression to the mean and serial dependence effects. We found significant differences in the accuracy of participants responses depending on whether the same physical stimuli were masked or not. We also found that participants' responses to the stimuli visible in both conditions were affected by the masking of the shortest and longest durations, suggesting that masked durations are not taken into consideration when computing summary statistics of stimulus distributions. Our prediction was that reports to the visible stimuli would be biased

towards the mean of the visible stimuli – participants would overestimate the duration of the stimuli when the shortest durations were masked and underestimate the duration of the stimuli when the longest durations were masked. However, we found the opposite effect – when we masked the longest durations, participants overestimated the duration of the visible stimuli, and when we masked the shortest durations, participants underestimated the duration of the visible stimuli.

Our data suggest that duration estimation requires conscious awareness of the stimulus which has some implications for theories on the mechanisms underlying the perception of time. Focusing specifically on the relevance of our findings for theories of time perception, scalar timing theory does not make explicit predictions regarding the role of conscious awareness in the working of the internal clock, but it can be predicted that awareness is not needed for the activity of either the pacemaker or accumulator. Some formulations of the internal clock model specify that attention has an effect on the activity of the switch component of the internal clock (Zakay & Block, 1996; Lejeune, 1998). However, attention does not require awareness (e.g., Hsieh, Colas, & Kanwisher, 2011), so even under these models, it would be possible that a system following scalar timing theory could form an estimate of duration without the stimulus reaching awareness. A human observer's awareness of the duration estimate would be required, but they do not need to necessarily be aware of the stimulus in order to build that estimate. In contrast, our findings that participants could not estimate the duration of masked stimuli provide evidence in favour of the theory that duration estimations are derived from the observers' perceptual content. Roseboom et al. (2019) suggest that perception of time arises from the perception of changes in perceptual content. Consequently, if an event does not reach conscious awareness, it will have no effect on perceptual content and its duration will not be possible to be estimated.

In addition, we also provide evidence that attractive perceptual biases in duration require conscious awareness. Previous studies have found that adaptation to numerosity can occur outside of awareness (Liu et al., 2013). In contrast, masked stimuli were found to contribute to regression to the mean but not serial dependence in an orientation estimation task (Suárez-Pinilla, 2018). Based on the available evidence it is difficult to conclude whether the difference between these studies and our results is due to differences in how the brain processes orientation, numerosity, and duration or differences between the three types of investigated biases – sensory adaptation, serial dependence, and regression to the mean. On the one hand, orientation is considered a low-level feature of visual stimuli that is encoded from the earliest instances of sensory processing in the brain and thus differs from both numerosity and

duration which can be considered as more abstract stimulus features (Whitney & Leib, 2018). It is commonly suggested that perception of duration and numerosity could be based on a common neuronal mechanism (Walsh, 2003). Based on this assumption, it is possible that the difference between our results and the results of Liu et al. (2013) could be pointing towards a fundamental difference between repulsive and attractive biases – sensory adaptation does not require conscious awareness while serial dependence and regression to the mean do. In support of this proposition, researchers have found multiple differences in the characteristics of sensory adaptation and serial dependence, which suggest that these opposing biases might occur at different stages of perception (Fisher & Whitney, 2014; Fritsche, Mostert, & de Lange, 2017; Suárez-Pinilla et al., 2018). However, this cannot completely explain the discrepancy between our findings and the findings of Suárez-Pinilla (2018). If masked orientations do not give rise to an attractive bias on a local scale, then it is unlikely that they would contribute to global attractive biases. Suárez-Pinilla (2018) suggest that their results could be explained by the relatively short lengths of the blocks and transfer of priors between conditions such that participants formed global rather than condition-specific priors. As discussed below, we suspect that participants in our experiment were also forming a global prior based on the duration estimation training conducted at the beginning of each session, and that prior was only partially shifted during the experimental task.

This study yielded two unexpected findings that require further consideration – we found only limited evidence for serial dependence in visible durations and, more globally, we found evidence of repulsion from rather than regression to the mean of the presented stimuli. Previous research has found that the direction of trial-to-trial local biases in perceptual decisions is related to confidence – trials associated with high confidence give rise to positive biases, while trials associated with low confidence give rise to negative biases (Suárez-Pinilla, 2018). Focusing on our experiment, it is possible due to the characteristics of the main task, duration judgements even of visible stimuli were not associated with very high confidence – participants received no feedback (this would have revealed the durations of the masked stimuli), and the stimuli were low contrast with gradual onset and offset (to make sure they did not break through the mask). This could explain why we see mixed evidence for serial dependence, and possibly why we see repulsion from the mean. At the beginning of every session, participants completed a duration estimation training task in which they had to estimate a wide range of durations (including the stimuli that would, later on, be masked in the main CFS task) and received feedback on their performance. It is, therefore, possible trials in the training task were associated with high confidence and as a result, gave rise to very

strong expectations which the main CFS task was unable to shift. As a result, participants could have been biased towards providing responses closer to the stimuli they were expecting to see but were missing. Evidence that the training task might have created very strong expectations comes from the finding that in both conditions participants responses to the masked stimuli were closer to the mean of all durations rather than the mean of only the visible stimuli.

To conclude, the current study provides initial evidence that estimation of stimulus duration requires conscious awareness, and as a result, durations presented outside of awareness probably do not generate perceptual biases. Nonetheless, we cannot draw strong conclusions on the basis of this data alone as our inability to show convincing serial dependence following visible durations suggests our analysis was underpowered, and it is not possible to determine if the repulsion from the mean effect we observed could stem from unmeasured effects of confidence or our task design.

Chapter 7. Effect of spontaneous blinking on performance in a duration bisection task

This final empirical chapter presents the preliminary findings of an experiment investigating the role of dopamine in time perception. We aimed to replicate three experiments conducted by Terhune et al. (2016) which showed that spontaneous blinking (a proxy measure of dopaminergic activity) leads to an overestimation of durations in a duration bisection task. While we have not completed data collection yet, our preliminary findings across two experiments (N=11 and N=9) do not provide support for Terhune et al.'s (2016) results, showing no conclusive evidence of an effect of blinking on performance across both sub-second and supra-second visual intervals. If the final dataset does support the prediction that spontaneous blinks affect responses in duration bisection, further experiments examining if the effect extends to other time perception tasks would be necessary to ascertain whether dopaminergic activity affects responses by shifting decision biases or by influencing the processing of temporal stimuli.

7.1 Introduction

There is an ongoing debate regarding the role of the dopaminergic system in time perception. Proponents of the scalar timing theory suggest that dopamine influences the speed of the internal clock (Meck, 1995). This proposition has been supported by pharmacological studies conducted on animals (Maricq & Church, 1983; Meck 1983), and human patients with dopamine-related disorders such as Parkinson's (Pastor et al., 1992), although the evidence for the latter is mixed (e.g., Riesen & Schnider, 2001; Smith et al., 2007).

Studies have suggested that the role of dopamine in time perception can also be investigated non-pharmacologically by using spontaneous blinking as a proxy measure for dopaminergic activity, as the rate of spontaneous blinking has been found to correlate with dopamine release (e.g., Karson, 1983). Using this technique Terhune et al. (2016) discovered that in a duration bisection task, stimuli presented following a blink were overestimated compared to stimuli preceded by no blinks. These results support the proposition that dopamine influences the rate of the pacemaker in the internal clock— as dopamine increases, the pacemaker generates more ticks for the same physical duration, leading to the overestimation of the presented interval and therefore increasing the propensity to classify stimuli as long rather than short. However, another study found no effect of spontaneous blinking on performance

in a duration estimation task (Suárez-Pinilla et al., 2019), raising concerns about the generalizability of Terhune et al.'s (2016) findings, the role of dopamine in time perception, and the reliability of spontaneous blinking as a measure of dopaminergic activity. For example, evidence suggests that dopaminergic activity can affect performance on perceptual tasks by inflicting response biases (Wang et al., 2018), or influencing motivation (Soares et al., 2016) rather than changing perception. In addition, studies on the relationship between spontaneous blinking and dopamine release have provided conflicting results (e.g. van der Post, de Waal, 2004), so it is unclear whether Suárez-Pinilla et al.'s (2019) findings reflect no relationship between dopamine and duration estimation, or no relationship between dopamine and spontaneous blinking. For a detailed review of the surrounding literature, please refer to Chapter 2 (*Neurophysiology of time: Dopamine*, p.39).

Here, we decided to replicate the study by Terhune et al. (2016) to directly investigate the reliability of their findings and re-examine whether their results reflect an influence of spontaneous blinking, as a proxy for dopamine activity. Following Terhune et al.'s (2016) experiment, participants completed a duration bisection task – they were presented with a range of durations and had to classify the stimulus presented on each trial as either long or short. If spontaneous blinking leads to an overestimation of durations, then the psychometric function fitted to participants' probability of long responses should exhibit a leftward shift – the bisection point (50% threshold) associated with post-blink-trials would be shorter than the bisection point associated with trials preceded by no blinks.

7.2 Method

7.2.1 Pre-registration and data availability

The hypotheses, design, experimental procedure, and analyses are pre-registered at the Open Science Framework (<https://osf.io/x9zc8>). The eye-tracking data and the behavioural data is available at: <https://osf.io/2p3fv/files/>.

7.2.2 Design

In this experiment participants had to complete a duration bisection task – participants were presented with stimuli of different durations and had to classify the duration presented on each trial as short or long. The dependent variable of interest was the ~50% threshold, or bisection point, of the psychometric function fitted to participants data from the bisection task.

The range and modality of the stimuli were manipulated between participants and had three levels – sub-second visual intervals (7 stimuli from 300 to 700ms in steps of 67ms), supra-second visual intervals (7 stimuli from 1400 to 2600ms in steps of 200ms), and sub-second auditory intervals (7 stimuli from 300 to 700ms in steps of 67ms). Here we presented results regarding conditions 1 and 2 – the sub-second and supra-second visual conditions.

Trials were additionally divided based on trial type into post-blink and post-no-blink trials. Blinks were defined using the autodetected blink events from the eye-tracking output (start of eye blink and end of eye blink). If at least one blink began or ended within the post-stimulus interval of the previous trial, the current trial was labelled as a post-blink trial. If on the previous trial, a participant kept their eyes open throughout the entire post-stimulus interval, the current trial was labelled as a post-no-blink trial. Post-stimulus intervals varied randomly between 800-1200ms in condition 1 (visual sub-second condition), 900-2100ms in condition 2 (visual supra-second condition), and was fixed to 1000ms in condition 3 (auditory sub-second condition).

7.2.3 Procedure

Participants completed three blocks in each condition. Each block was divided into 1 training run and 3 testing runs. The training run contained 10 repetitions of the shortest and longest duration pertaining to the appropriate condition, and participants' task was to classify whether the stimulus on each trial was the short or the long stimulus. At the end of the trial, participants were provided with feedback related to the accuracy of their responses. During the testing runs, participants were presented with a range of durations between the shortest and longest anchor from the training run, and participants' task was to determine if the stimulus presented on each trial appeared to be more similar in duration to the short or long anchor. As this was an appearance task, participants were not provided with feedback. In conditions 1 and 3 each stimulus was presented on 90 trials overall, and in condition 2 each stimulus was presented on 72 trials overall.

Every trial started with a presentation of a fixation cross which varied between 1250ms and 1450ms. After that, the stimulus was presented, followed by a post-stimulus interval that varied from 800ms to 1200ms in condition 1, and from 900ms to 2100ms in condition 2. After that, a response cue was presented prompting participants to classify the presented interval as short or long by pressing a key on the keyboard, J or N, with the mapping between key and response counterbalanced between participants (Figure 7.1). In the training blocks, following

the participant's response, feedback indicating whether their response was correct or not was presented for 1000ms. Instructions can be found in Appendix J.

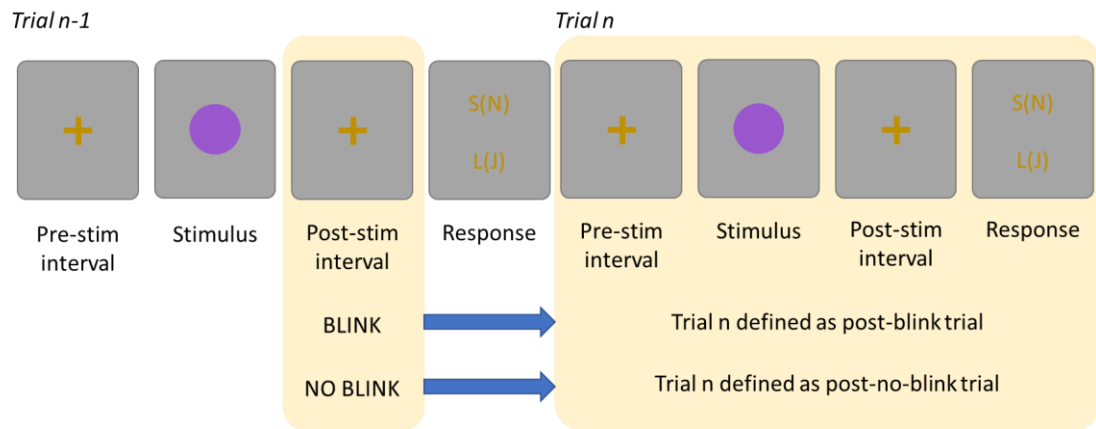


Figure 7.1. Trial structure. *On every trial in the testing runs, participants had to classify the presented stimulus as either more similar to the short anchor or to the long anchor. If the participant blinked during the post-stimulus interval of the previous trial, the current trial was marked as a post-blink trial. If the participant did not blink during the post-stimulus interval of the previous trial, the current trial was marked as a post-no-blink trial.*

7.2.4 Apparatus

The experiment was run on Matlab2018b using Psychtoolbox-3.0.14 (Brainard, 1997, Pelli, 1997; Kleiner, Brainard, Pelli 2007) and the Eyelink toolbox (Cornelissen, Peters, & Palmer, 2002). An Eyelink 1000 Plus eye tracker with 500Hz sampling rate was used to detect eye blinks from participants' right eye.

The visual stimulus consisted of a purple circle, 13dva, presented on a grey background. The stimulus was calibrated to have the same luminance as the background. The fixation cross and all written instructions presented during the experiment were presented in dark yellow and also matched the luminance of the background. The experiment was presented on a LaCie, Electron 22 Blue II 22-inch monitor with a resolution of 1024x768 pixels and 120Hz refresh rate.

Participants sat at 57cm away from the screen and rested their head on the chinrest throughout the entire duration of a block. The eye tracker was calibrated at the beginning of every block using a 5-point grid. The experiment was started only if the average error recorded during the calibration was below 0.5 degrees of visual angle. One participant for whom the eye tracker could not be calibrated did not take part in the experiment.

7.2.6 Analyses

Using the Palamedes toolbox (Prins, & Kingdon, 2009) for Matlab, we fitted logistic psychometric functions to the proportion of long responses each participant gave to the stimuli presented on trials preceded by blinks and trial preceded by no blinks separately. Replicating the analysis detailed by Terhune et al. (2016), the threshold and slope parameters of the psychometric function were free to vary, the guess rate was fixed to 0, and the lapse rate was fixed to 0.1. For each participant, we extracted the 50% thresholds, or bisection points, from the functions fitted to the post-blink and post-no-blink trials, and we used Bayesian t-tests to investigate if there was a difference in the bisection points arising from post-blink and post-no-blink trials. The expected effects for the Bayesian t-tests were based on the results reported by Terhune et al. (2016) – 44ms for condition 1, and 69ms for condition 2. Finally, for each trial type, we computed bias-corrected and accelerated bootstrap intervals (Efron, 1987) across participants. We carried out an additional analysis where all parameters of the logistic function were free to vary. The rest of this analysis followed the same steps as above.

7.2.5 Exclusion criteria

We followed the exclusion criteria detailed by Terhune et al. (2016). Participants were excluded if they blinked during the post-stimulus interval on less than 10% of the trials. Participants were also excluded if the logistic function fitted to either the post-blink or post-no-blink trials showed poor fit to the data. Goodness-of-fit was assessed using the Palamedes toolbox (Prins, & Kingdon, 2009; PAL_PFML_GoodnessOfFit function with 1000 simulations), and participants with pDev values smaller than 0.05 were excluded from the analysis. We also excluded participants if the difference in bisection point for post-blink or post-no-blink trials was more than 3 standard deviations away from the mean bisection point for the respective trial type averaged across all participants.

7.2.7 Participants

The required sample sizes were determined using Bayesian t-tests. For each condition, we determined the smallest sample size that would yield a Bayes factor greater than 3 or smaller than 1/3 if we expect to find results similar to those reported by Terhune et al. (2016). The prior expected effects and observed effects in the t-tests were set to the differences in bisection points between trials preceded by blinks and trials preceded by no blinks reported by Terhune et al. (2016): 44ms for the sub-second visual condition (condition 1), and 69ms for the

supra-second visual condition (condition 2). The standard deviations of the observed differences were calculated based on the results provided by the authors: 82.64ms for condition 1, and 146.05ms for condition 2. The sample sizes with which we can expect to obtain a Bayes factor greater than 3 are 14 participants in condition 1, and 17 participants in condition 2. When we collect these initial samples, we will conduct the analyses detailed above. If we have not exceeded our sensitivity threshold (BF_{10} smaller than 1/3 or BF_{10} greater than 3), we will continue data collection and reanalyse the data after every participant until we reach a maximum of 30 participants per experiment. We have not reached the minimal sample size yet, so the results presented in this paper are only preliminary.

A total of 41 participants (N women = 28; M_{age} = 22.38 years, SD_{age} =6.42) took part in the experiment. Of them, 19 participants had a full set of eye-tracking data and corresponding behavioural data for condition 1, and 14 for condition 2. 2 participants from condition 1 and 1 from condition 2 were excluded due to insufficient blinks. In the analysis replicating Terhune et al. (2016), 6 participants were excluded due to poor model fit ($pDev < 0.05$) from both condition 1 and 2. One participant from condition 1 was excluded due to outlying bisection point values ($> Mean_{BP} + 3SDs$), no outliers were found in the data corresponding to condition 2. In the additional analysis detailed above, 2 and 5 participants were excluded due to poor model fit ($pDev < 0.05$) from condition 1 and 2 respectively, and 1 participant from condition 1 was excluded due to outlying bisection point values ($> Mean_{BP} + 3SDs$).

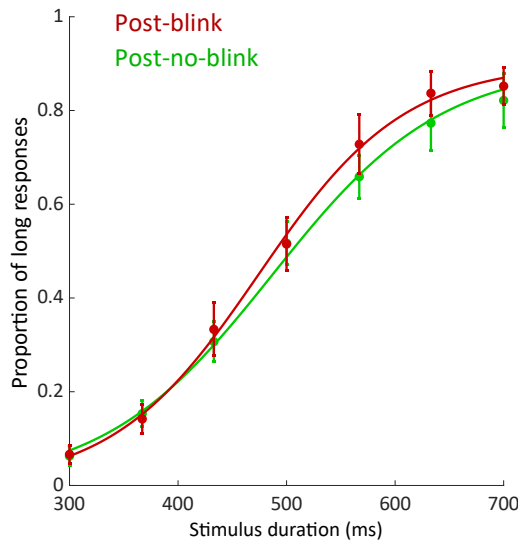
To summarise, the analysis replicating Terhune et al. (2016) was based on 11 participants in condition 1, and 7 participants in condition 2. The additional analysis was based on 15 participants in condition 1 and 9 participants in condition 2.

7.3 Results

First, we replicated the analysis detailed by Terhune et al. (2016). On average, participants responses in the two conditions did not show a consistent difference between post-blink and post-no-blink trials (Figure 7.2). In the visual sub-second condition, the bisection point in post-no-blink trials, M_{BP} = 480.46ms (SD =40.40), was greater than the bisection point in post-blink trials, M_{BP} = 467.15 (SD =51.56), indicating that durations presented following a blink appeared longer. However, the mean difference we found, $Mean_{diff}$ = 13.31ms, bootstrap CI95%: [1.80. 25.55], was smaller than the result obtained by Terhune et al. (2016) providing inconclusive evidence in favour of the alternative hypothesis, BF_{10} = 1.52. In the visual supra-second condition, there was no difference in the bisection point associated with post-no-blink trials, M_{BP} = 1955.42ms, (SD = 141.52), and post-blink-trials, M_{BP} = 1955.67 (SD =88.85), $Mean_{diff}$ = -

0.25ms, bootstrap CI95% [-61.69, 57.97], providing inconclusive evidence in favour of the null hypothesis: $BF_{10} = 0.62$.

A. Visual sub-second condition



B. Visual supra-second condition

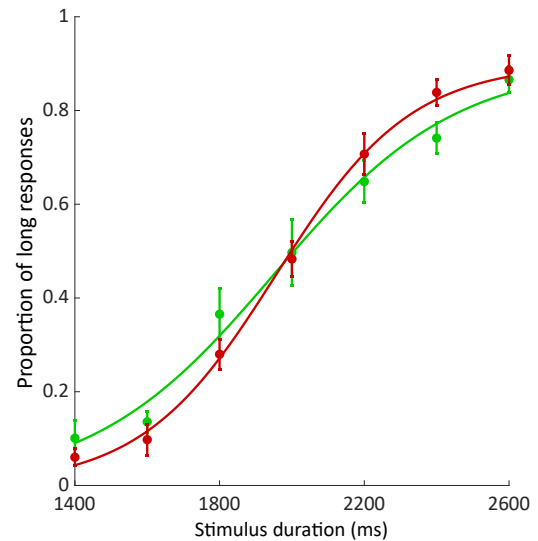


Figure 7.2. Behavioural results. Average proportion of long responses for each stimulus on post-blink (red), and post-no-blink trials (green) in the sub-second visual condition (A), and the supra-second visual condition (B). These are the results of the participants included in the analysis replicating Terhune et al. (2016). Error bars show $\pm 1SE$. The solid lines in both figures show the psychometric curves fitted to the average data of all participants.

The additional analysis we carried out provided similar results. The difference in bisection points between post-no-blink and post-blink-trials in condition 1, $Mean_{diff} = 2.24$, bootstrap CI95% [-24.57, 22.055], provided inconclusive evidence in favour of the null hypothesis: $BF_{10} = 0.34$. The difference in bisection points in condition 2, $Mean_{diff} = -3.44$, bootstrap CI95% [-62.67, 45.42], also provided inconclusive evidence in favour of the null hypothesis: $BF_{10} = 0.46$.

7.4 Discussion

The experiment presented here aims to investigate if spontaneous blinking, as a proxy for dopaminergic activity, leads to an overestimation of durations. Proponents of scalar timing theory predict that an increase in dopamine levels speeds up the internal clock leading to overestimation of durations. To date, only two studies have investigated the effect of spontaneous blinking on performance in time perception tasks, and their results are contradictory. Terhune et al. (2016) found that blinks affected performance on a duration bisection task, while Suárez-Pinilla et al. (2019) found no effect of blinking on duration estimation of naturalistic stimuli. The current study aimed to replicate the experiment

conducted by Terhune et al. (2016). Participants completed a duration bisection task, and we tested the hypothesis that the stimulus duration which participants classify as long on 50% of the trials would be longer if a blink occurred during the previous trial compared to if a blink did not occur the previous trial. In the data we have obtained to date, we found conclusive evidence in favour of neither the alternative nor the null hypothesis. We will need to finish data collection to be able to provide stronger conclusions, but our current results suggest that if an effect of spontaneous blinking on duration estimation does exist, the effect will be likely smaller than the effect reported by Terhune et al. (2016).

Even though we do not have the complete results for this experiment yet, we can discuss the possible implications of failing or succeeding to replicate the findings of Terhune et al. (2016). To begin with, even if we find evidence of no effect of spontaneous blinking on performance in the duration bisection task, this will not necessarily suggest that dopamine has no role in time perception. Instead, it is possible that spontaneous blinking is not a reliable proxy measure for dopaminergic activity, which is supported by previous studies investigating the relationship between dopamine and blinking in humans (van der Post et al., 20114). For example, it was found that administration of a dopamine antagonist and a dopamine agonist had no effect on spontaneous blinking in healthy adults (van der Post et al., 20114; Dang et al., 2017) and that there was no relationship between spontaneous blink rate and the availability of dopamine receptors (D2 receptors in particular; Dang et al., 2017). Furthermore, a dopamine antagonist was reported to *increase* spontaneous blink rate in non-human animals (Kotani et al., 2016). Finally, another study found that the effect of a dopamine agonist on spontaneous blink rate depended on the participants' baseline blink rate. Participants with low blink rate at baseline exhibited an increase in blink rate following the application of the agonist, while participants with high blink rate exhibited a decrease in blink rate following the application of the agonist (Cavanagh, Masters, Bath, & Frank, 2014). Combining the results of these studies, there is still a lot that is unknown about the relationship between the dopaminergic system and spontaneous blinking, suggesting that spontaneous blinking is, for now, not a reliable measure of dopamine activity.

Another possible outcome of this experiment after completing data collection is that our results replicate the findings reported by Terhune et al. (2016) – the bisection point on trials following blinks is shorter than the bisection point on trials following no blinks. However, a difference in bisection points alone is not enough to disentangle whether dopamine's role is specific to the duration estimation component of the task or is related to response or decision biases that are independent of the duration estimation process (Figure 7.3A). The bisection

point associated with post-blink trials will also be shorter if an increase in dopaminergic activity is associated with a general increase in the probability of long responses across the whole range of tested stimuli (Figure 7.3B1). The analysis performed by Terhune et al. (2016) that we are replicating here cannot differentiate between these two possibilities, and one should also consider the participants' responses across all presented stimuli. For example, previous studies have found that increase in dopaminergic activity increases the proportion of positive responses in a change detection task (Wang et al., 2018), while a decrease in dopaminergic activity results in a decreased rate of responses in duration estimation task (Soares et al., 2016). Changes specific to participants' bisection point in the absence of a general increase in long responses (Figure 7.3B2) also have two potential explanations. A horizontal shift in the psychometric function could indicate a shift in perceived duration (Figure 7.3C2), but also a shift in criterion – when dopamine is higher, observers may shift their criterion for replying “long” towards shorter durations (Figure 7.3C1). No study to date has managed to successfully disentangle these possibilities. Therefore, to date, it is still not clear to what extent dopamine's role in duration estimation might be related to the timing component of the task.

If Terhune et al.'s (2016) results replicate, further research would be necessary to find out whether this effect generalizes across other duration estimation tasks in order to improve our understanding of the specific role that dopamine plays in time perception. One possibility would be to again investigate the effect of blinking on duration estimation, but using simpler, more controlled stimuli than the naturalistic videos used by Suárez-Pinilla et al. (2019), which could have altered motivation and attention on a trial-by-trial basis thus introducing confounds into the design. Duration bisection and duration estimation tasks differ in the way in which decisions are reached, and responses are provided. For example, in a duration bisection task, it is enough to have a single criterion separating long from short responses – a decision can be reached based on whether the sensory evidence surpasses that criterion or not. In contrast, a duration estimation task would be associated with multiple interrelated decision criteria, as there are multiple possible responses. Therefore, if spontaneous blinking leads to an overestimation of durations across both duration bisection and duration estimation tasks, this could provide strong evidence for a role of dopamine in time estimation rather than decision biases.

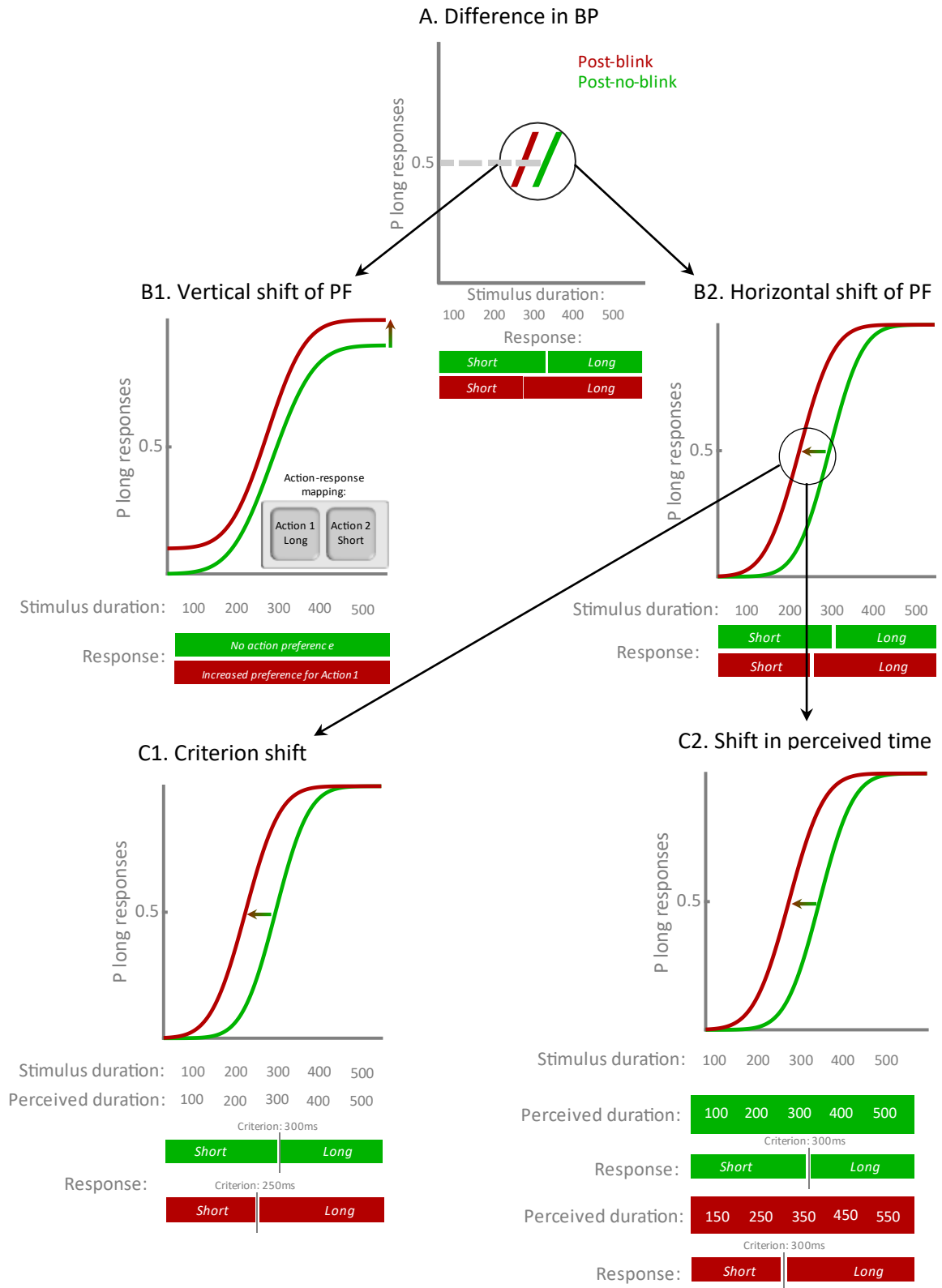


Figure 7.3. Possible effect of spontaneous blinking on duration bisection. (A) Analysis limited to difference in bisection points between post-blink and post-no-blink trials cannot distinguish between vertical (B1) and horizontal (B2) shifts in the psychometric function. If spontaneous blinking results in a horizontal shift of the psychometric function (B2), it would be difficult to distinguish whether this change is due to a shift in participants' response criterion for reporting long (C1) or due to a shift in participants' subjective mapping of durations (C2).

Chapter 8. General discussion

Here I discuss how the empirical work presented in the previous chapters informs the existing literature on perceptual biases in duration estimation¹, I provide some suggestions for reconciling the unexpected findings in some of the reported studies, and suggest some further directions of how we plan to build on the findings reported here.

In addition, while conducting the research presented here, we became aware of some issues that previous studies investigating regression to the mean in duration perception either have not reported or have not taken into account. The concerns we raise are also relevant to the wider field of time perception research in general, but here I focus specifically on the studies investigating Bayesian perceptual biases. The first issue we discovered was that most participants we recruited were not able to do duration reproduction to a degree that indicates they could accurately perceive the presented stimuli. This raises concerns regarding the generalizability of any effect based on such tasks. Previous studies also do not reliably report concrete exclusion criteria determined prior to data collection to ensure that the data is suitable for drawing conclusions regarding the mechanisms of human time perception. For reference, this is the equivalent of investigating the processes underlying colour perception without excluding participants who are colour-blind. Furthermore, if a task is based on the assumption that most people are not colour-blind, but in fact the opposite is true, then the findings will apply only to a small group of people rather than to the general population overall. In addition, most studies do not report taking any special precautions to ensure the timing accuracy of the equipment used to record responses raising concerns about the reliability of the data used for analyses.

8.1 Summary of key findings

In Chapter 4 we found evidence that when other distributional statistics are kept constant, the variance of the prior distribution influences the magnitude of regression to the mean effects – a peaked (Gaussian) stimulus distribution gave rise to a stronger bias towards the prior than a wide (uniform) distribution with the same range and mean. Our findings also support and extend the results of previous studies which report that simultaneous changes in the range and variance of stimulus distributions – stimulus distributions with a narrow range and

¹ The findings of Chapter 7 are not discussed further because the preliminary findings I presented do not allow us to draw conclusions strong enough to contribute to the overall literature on the role of dopamine in time perception and Bayesian inference.

variance were found to general stronger biases than stimulus distribution with a wide range and variance (Körding et al., 2004; Miyazaki et al., 2005; Berniker et al. 2010; Garrido et al., 2013). To our knowledge, this is the first study to investigate and find support for the specific prediction of Bayesian brain theories that difference in the variance of the prior distribution alone modulates the effect of prior expectations on perception.

In Chapter 4, we also conducted a replication of an experiment by Acerbi et al. (2012) where we investigated the effect of variance in peaked skewed distribution compared to a uniform distribution. The skewed distribution was constructed by increasing the presentation frequency of the second shortest stimulus in the range relative to the other stimuli in the range. We found that the skewed distribution with small variance did not give rise to a stronger regression to the mean effect compared to the uniform distribution. On average, participants' behavioural responses diverged from the predictions of the Bayesian model, showing a decrease instead of an increase in the magnitude of the bias for stimuli not in the immediate vicinity of the overrepresented stimulus. In addition, as the skewed distribution had a shorter mean than the uniform distribution, we predicted that participants' responses would be, on average, shorter in the skewed condition. Although we found that in the skewed condition participants were biased towards a shorter value compared to the uniform condition, the difference we found was smaller than the effect reported by Acerbi et al. (2012), and we did not manage to replicate their results.

In Chapter 5, we successfully replicated the range effect of regression to the mean in a duration reproduction and a duration estimation task – the same physical duration was overestimated when it was shorter than the mean but underestimated when it was longer than the mean. These results provide further support that context effects are ubiquitous in perception (e.g., Jazayeri & Shadlen, 2010; Petzschner & Glasauer, 2011; Petzschner et al., 2015). Furthermore, we found evidence of potential EEG correlates of this context effect. Most notably, an MVPA classifier was able to distinguish stimulus context based on the EEG activity recorded for the same physical stimulus. Damsma et al. (2020) recently reported similar findings, and together, our results strongly suggest that brain activity reflects not only the physical characteristics of the presented stimuli but also participants' subjective experience which is affected by their prior expectations.

Nonetheless, the main aim of the research presented in Chapter 5 was to investigate the effect of sensory precision on perceptual biases. However, we did not find conclusive evidence in favour of the prediction of Bayesian models that sensory precision should modulate the effect

of prior expectations on perception. We found no difference in regression to the mean in duration reproduction between participants with high and low sensory precision, and we found a much weaker correlation between sensory precision and regression to the mean than the correlation predicted by a Bayesian model. Finally, we found no difference in the ability of the MVPA classifier to distinguish stimulus context between participants with high and low sensory precision. To date, very few studies have investigated the relationship between sensory precision and regression to the mean, and our results contradict their findings. For example, Cicchini et al. (2012) found that regression to the mean and sensory precision were positively correlated. The researchers used different measures to estimate these parameters, and as already discussed in Chapter 5 (p.113-114) it is possible that we did not find the effects we were expecting because the estimates we used do not reflect participants' true sensory precision and/or regression to the mean. An alternative possibility is that regression to the mean arises from a different process, not entirely consistent with the Bayesian models we have examined. This is discussed further below.

Finally, the results of Chapter 6 suggest that durations could not be estimated outside of awareness and that stimuli presented outside of awareness did not contribute to global and local attractive biases (regression to the mean and serial dependence respectively). A recent study on the role of conscious awareness in perceptual biases in orientation perception provides mixed support for our findings (Suárez-Pinilla, 2018). On the one hand, serial dependence was found for visible but not masked stimuli. On the other, it was also found that estimates of mean orientation were based on both visible and masked stimuli, while in our study, masked stimuli did not contribute to regression to the mean. As discussed in Chapter 6 (p.127), it is likely that the discrepancy between these findings arose due to the experimental design of the study. In addition, a study by Liu et al. (2013) found that sensory adaptation to numerosity can occur when stimuli are masked. Our findings that masked durations do not give rise to attractive biases can be interpreted as providing evidence against a common mechanism underlying magnitude estimation across domains (Walsh, 2003). Alternatively, it is possible that repulsive and attractive biases occur at different stages of processing, and while adaptation does not require awareness, regression to the mean does. This proposition fits with recent propositions that adaptation occurs at the stage of sensory processing, while serial dependence arises at the stage of decision making (Fritsche et al., 2017).

8.2 Wider implications

From the findings summarized above, two results that were particularly unexpected given the predictions of the Bayesian brain hypothesis and the specific Bayesian ideal-observer models that we used to form our specific predictions – first, that the skewed peaked distribution gave rise to a mixture of negative and positive biases, and second, that sensory precision was not related to regression to the mean. Here I present some speculations on how these issues could be resolved. Considering the increased interest in open science across neuroscience and psychology, I also discuss some issues with transparency and data quality in the literature on duration reproduction.

8.2.1 Reconciling attractive and repulsive perceptual biases

Most of the research I have presented here focuses on biases towards the prior – serial dependence on a local level, and regression to the mean on a global level. However, in Chapter 4 we found that a peaked, skewed stimulus distribution gave rise to a mixture of attractive and repulsive biases. Repulsive biases in perception have often been accounted for by sensory adaptation (Webster, 2011), and typically arise when the repeated presentation of a particular stimulus value causes the perception of other subsequently presented stimulus values to be repelled from it (Webster, 2011). It has been proposed that this perceptual effect results from a change in neuronal excitability and selectivity – the repeated presentation of a stimulus causes a decrease in the firing rate of neurons selective for the adapted value (Kohn, 2007).

While it may seem contradictory that prior experiences can lead to both attractive and repulsive biases, researchers have found differences in the experimental conditions that give rise to neuronal adaptation and serial dependence. For example, it was recently found that the direction of perceptual biases changes with trial history – the strongest serial dependence effect was induced by the most recently presented stimulus, while stimuli presented further back into the trial history gave rise to negative biases (Suárez-Pinilla et al., 2018). The same study also found that trials associated with high decision confidence gave rise to serial dependence, while low confidence trials gave rise to adaptation effects. Other studies report evidence that serial dependence in vision still occurs when subsequent stimuli are presented in different locations on the retina (Fisher & Whitney, 2014), while sensory adaptation in vision is retinotopic (Wenderoth & Wiese, 2008; Morgan, 2014). Finally, adaptation to an abstract stimulus feature, i.e. numerosity, was found to occur outside of awareness (Liu et al., 2013), while attractive biases to an abstract stimulus feature, i.e. duration, were not found for masked stimuli (see Chapter 6). The differences between adaptation and serial dependence

suggest that these processes might be operating at different levels of perception – adaptation evolves at the stage of sensory processing while serial dependence (and positive biases in general) evolve later, probably at the stage of decision making (Fritsche et al., 2017).

Further to the empirically established differences between the effects, researchers have also proposed that repulsive and attractive biases may have different functional roles, suggesting that it could be possible to reconcile these opposing biases within a single perceptual framework. On the one hand, following adaptation, subjective reports of stimulus values close to the adaptor are repulsed away, suggesting that stimulus discriminability has increased – observers become more sensitive to stimulus changes close to the adapted value. As a result, researchers have proposed that adaptation serves to improve observers' ability to detect changes in the environment (Clifford & Wenderoth, 1999; Kohn, 2007; Webster & MacLeod, 2011; Webster 2015; Fritsche et al., 2017). On the other hand, researchers propose that attractive biases help stabilize perception over short periods of time (Fisher & Whitney, 2014; Fritsche et al., 2017; Cicchini, Mikellidou, & Burr, 2018). If we assume that the state of the external world remains relatively stable over short timescales, then a positive bias towards previous experience would smooth out perceptual errors and (hopefully) result in optimal decision making. The distinction in the potential roles of sensory adaptation and attractive biases further fits with the proposition that they operate at different stages of perception – adaptation is inducing changes in sensory sensitivity during initial stimulus processing, while serial dependence and regression to the mean are influencing decisions to optimize behaviour.

Probabilistic perceptual models based on different mechanisms have been proposed to explain attractive and repulsive perceptual biases. First, it has been proposed that Bayesian inference provides the only available domain-general mechanism for conceptualizing the effect of prior expectations on behavioural reports observed in magnitude estimation tasks across modalities (Petzschner et al. 2015). Furthermore, Petzschner and colleagues (Petzschner & Glasauer, 2011; Petzschner et al., 2015) show that Bayesian models can resolve the apparent conflict between classical psychophysical laws. Both the Weber-Fechner law and Stevens' power law hold if we assume that the brain represents the sensory likelihood and prior in logarithmic space, and transforms the posterior distribution into linear space for the purpose of providing a response. Therefore, the Bayesian model detailed by Petzschner et al. (2015) constitutes a powerful framework for investigating the role of prior expectations on decisions in any task based on magnitude estimation.

Sensory adaptation in duration estimation has been accounted for by a probabilistic model based on population codes. Heron et al. (2012) suggest that time perception is based on the activity of a population of duration-selective channels. Each channel has a preferred duration, and the channel's probability of response depends on how close the presented duration is to its preferred duration. In this model, the probability distribution of perceived duration results from the combination of the response probability distribution of a population of such duration-selective channels. Heron et al. (2012) successfully used this framework to explain adaptation effects in perceived duration – sensory adaptation was modelled as a reduction in the channel's response probability, with the magnitude of response reduction depending on the similarity between the adaptor and the channel's preferred duration. Similar to the generalizability of Petzschner et al.'s (2015) Bayesian ideal-observer model, the strength of the model proposed by Heron et al.'s (2012) is that it is based on a general brain-architecture principle (e.g., Hubel & Wiesel, 1968) that can explain sensory and neuronal adaptation across perceptual domains.

While Bayesian inference and the duration channel accounts provide good explanations for attractive and repulsive biases respectively, neither model can naturally account for the opposite bias². Therefore, models built to explain directional biases would fail to capture behavioural patterns that reflect both types of biases. In Chapter 4, we found that in a peaked distribution of durations, stimuli close to the over-represented stimulus show positive bias, while more distant stimulus values show a repulsive bias. Similarly, Fritsche et al. (2017) and Bliss et al. (2017) found that serial dependence was induced when the previous and current trial were relatively similar, while large changes in stimulus values between trials could lead to negative biases. Finally, as already discussed, Suárez-Pinilla et al. (2018) found that serial dependence and sensory adaptation coexist within the same task. Such findings suggest that it would be sub-optimal for researchers to continue to study regression to the mean and sensory adaptation in isolation.

One possibility to reconcile these findings would be to combine Heron et al.'s (2012) and Petzschner et al.'s (2015) models into a single hierarchical framework where population codes give rise to the sensory likelihood, and Bayesian inference is used to integrate the sensory likelihood with prior expectations. Thus, adaptation would occur as part of the sensory

² Bayesian models can account for sensory adaptation if we relax the assumption of Gaussianity. For example, Stocker and Simoncelli (2006a) proposed that adaptation may cause different changes in the shape of the sensory likelihood close and far from the adapted stimulus value (also see Wei & Stocker, 2015).

processing of the stimulus, and positive biases would occur at the time of decisions. Such a hierarchical model was recently put into practice (Suárez-Pinilla, 2018), and it was shown that it could predict both attractive and repulsive biases in the variance of motion stimuli. As both population codes and Bayesian inference provide domain-general accounts of perceptual biases, it is very likely that the hierarchical model proposed by Suárez-Pinilla (2018) would also account for opposing biases in other modalities such as duration perception.

8.2.2 Role of sensory precision and working memory in attractive perceptual biases

In Chapter 5, I presented two experiments in which we did not find conclusive evidence in favour of a relationship between sensory precision and regression to the mean. An exploratory analysis we conducted on the data collected in Chapter 4, comparing the magnitude of regression to the mean between musicians and non-musicians, casts further doubt on the potential effect of sensory precision on perceptual biases. Previous studies have reported that musicians tend to show weaker biases in duration estimation tasks (Cicchini et al. 2012; Aagten-Murphy et al. 2014), and also have smaller Weber fractions compared to non-musicians (Grondin & Killeen, 2009). Although investigating the effect of musical expertise on time perception was not the main goal of the duration reproduction experiment presented in Chapter 4, we collected basic data regarding participants' musical experience. Thirty-two participants in our sample played a musical instrument on a regular basis (at least once a week). The other 14 participants either had no musical experience or had not played an instrument for at least 5 years. We compared the Spearman correlation participants achieved at the end of the first full block they completed (576 trials). Contrary to the findings of previous studies, we found that the musicians, $\text{MeanSC} = 0.36$ ($\text{SD} = 0.22$), in our sample did not outperform the non-musicians, $\text{MeanSC} = 0.44$ ($\text{SD} = 0.18$). This was also the case when we specifically compared performance only between musicians ($N=15$) and non-musicians ($N=6$) who in their first session took part in the uniform condition: $\text{MeanSC}_{\text{musicians}} = 0.38$ ($\text{SD}=0.19$), $\text{MeanSC}_{\text{non-musicians}} = 0.44$ ($\text{SD} = 0.21$). While this experiment was not set up to test the effect of musical experience on perceptual biases, these findings are in stark contrast with the results of previous studies and provide further evidence that the relationship between sensory precision and regression to the mean may not be as straightforward as predicted by Bayesian ideal-observer models.

There are a number of possibilities that could explain why in Chapter 5, we did not find a relationship between sensory precision and regression to the mean that is consistent with the

predictions of Bayesian models. As discussed in the chapter-specific discussion (see p.113) it is possible that the measures we used did not accurately reflect participants' true sensory precision and perceptual bias in the duration reproduction and estimation tasks. An alternative possibility is that the pattern of responses we observed in the duration reproduction and estimation tasks was generated by a non-Bayesian process and that the magnitude of regression to the mean was therefore not related to sensory precision.

A key assumption of Bayesian brain theories is that subjective priors are updated iteratively as new information is encountered. Using Kalman filters to simulate updating priors, previous studies have shown that global attractive biases can be well modelled based on the accumulation of local, trial-by-trial biases (e.g., Petzschner et al., 2015). These local biases, referred to as serial dependence, have been widely studied (e.g., Fisher & Whitney, 2014; Liberman et al., 2014; Bliss et al., 2017; Czoschke, Fischer, Beitner, Kaiser, & Bledowski, 2019). Interestingly, there is a key difference in the general approach studies that focus on serial dependence take to model local attractive biases and the approach studies that focus on regression to the mean take to model global biases (Jazayeri & Shadlen, 2010; Acerbi et al., 2012; Petzschner et al., 2015). Namely, research on serial dependence suggests that local attractive biases could arise from processes related to working memory (Kiyonaga, Scimeca, Bliss, & Whitney, 2017). If we assume that global attractive biases arise from the summation of local biases, then the role of working memory on regression to the mean needs to be better understood.

Working memory is a system that allows information to be stored for a limited time while it is in use. The main evidence in favour of the proposition that serial dependence originates from working memory comes from a range of studies that report that an increase in the delay between stimuli and responses leads to an increase in the magnitude of serial dependence (Papadimitriou, Ferdoash, & Snyder, 2015; Bliss et al., 2017; Fritsche et al., 2017). As an increase in the delay between stimulus and response leads to an increase in the involvement of working memory (e.g., Bliss et al., 2017), these results suggest that the integration of the current and previous trial depends, at least partially, on working memory. If serial dependence was purely perceptual in nature, manipulating the delay between stimulus and response should not affect the magnitude of the bias exerted by the stimulus on the previous trial. However, this is not the case because studies show that serial dependence evolves as the current stimulus is held longer in working memory.

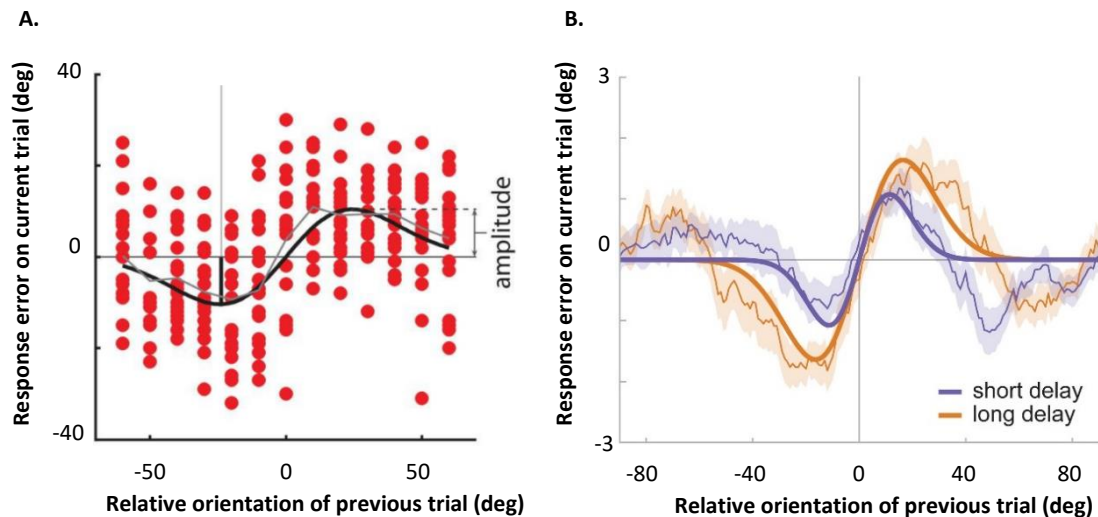


Figure 8.1. Using the derivative of the Gaussian (DoG) curve to model serial dependence. (A) shows the responses of one participant in an orientation estimation task modelled using the DoG curve (thick black line, figure adapted from Fisher & Whitney, 2014). The peak of the DoG curve reflects the strength of the serial dependence effect. (B) shows average serial dependence in an orientation estimation task (thin lines) modelled using the DoG Curve (thick lines). The serial dependence effect is stronger when there is a delay between stimulus and response (orange) compared to when responses are given immediately after stimulus offset (purple). The figure is adapted from Fritsche et al. (2017). Note that in both figures the curves are flipped compared to the example curves in Figure 8.2 because here the Y-axis shows the error between the response and the stimulus presented on the current trial rather than directly the provided response.

In the serial dependence literature, researchers use the first derivative of the Gaussian (DoG) curve to model attractive biases towards the stimulus presented in the previous trial (e.g., Fischer & Whitney, 2014; Bliss et al., 2017; Fritsche et al., 2017). It is worth noting that Bayesian ideal-observers models are process models based on pre-specified constructs, while models based on the DoG curve are purely descriptive. In that sense, when I refer to the DoG curve as a model, I refer only to the way it fits the data. As shown in Figure 8.1A the serial dependence effect is strong for stimuli that are close in value to the stimulus presented in the previous trial and diminishes as the difference between the previous and current stimulus increases, and the DoG curve fits this pattern of results well. Bliss et al. (2017) found that the spread of serial dependence, or the width of the DoG curve, varied across participants. Notably, in some participants, stimuli on the previous trial that were very different from the stimuli presented on the current trial even generated repulsive biases. Differences in the width of the DoG curve could be related to individual differences in working memory. As discussed above, the strength of the serial dependence effect depends on the involvement of working memory – the longer the current stimulus is kept in working memory, the stronger and more spread out the effect of serial dependence (Figure 8.1B, Fritsche et al., 2017). Therefore, the

differences in serial dependence Bliss et al. (2017) observed across participants could be explained by individual differences in working memory – participants who engage working memory more or have greater working memory capacity would exhibit stronger serial dependence. If there is an overlap in the processes that give rise to local and global attractive biases, then the DoG curve could also explain regression to the mean effects. The top row of Figure 8.2 provides an illustration showing how a DoG process could account for the overall context effect in regression to the mean and provide very similar results to a Bayesian process (figures A and B). In this illustration, the presented durations range from 200-450ms and 450-700ms. Depending on the width of the DoG curve, the two models could predict very similar responses for the presented stimuli and a very similar effect of context on the overlapping stimulus. We could expect such results to occur if the tested stimulus ranges are particularly narrow, such as in this example. However, if, for example, the range of the tested stimuli increases, the predictions of the two models begin to diverge. While most Bayesian models would predict that the regression to the mean effect would continue to increase³, a DoG model would predict that the attractive bias would decrease and eventually disappear.

If a DoG process can account for regression to the mean effects, then, individual differences in the magnitude of the bias could also be, at least partially, accounted for by individual differences in working memory, which is reflected in the width of the DoG curve. Figure 8.2C shows the effect of sensory precision on regression to the mean predicted by Bayesian models, while Figure 8.2D illustrates the possible effect of working memory on regression to the mean. The key difference here is that a Bayesian model predicts that high regression to the mean is related to poor sensory precision, while a DoG model predicts that high regression to the mean is related to good cognitive functioning. It is, therefore, possible that in Chapter 5 we were not able to predict the magnitude of regression to the mean across participants because the models we used are based on the presumption that regression to the mean is a perceptual effect based on sensory precision, rather than a mnemonic effect based on working memory. To date, the role of working memory in regression to the mean, and more generally in Bayesian perceptual inference, remains unknown, and further research will be necessary to establish the connections between these fields of study.

³ In some formulations, Bayesian models predict that eventually the regression to the mean reaches a plateau (e.g., Jazayeri & Shadlen, 2010).

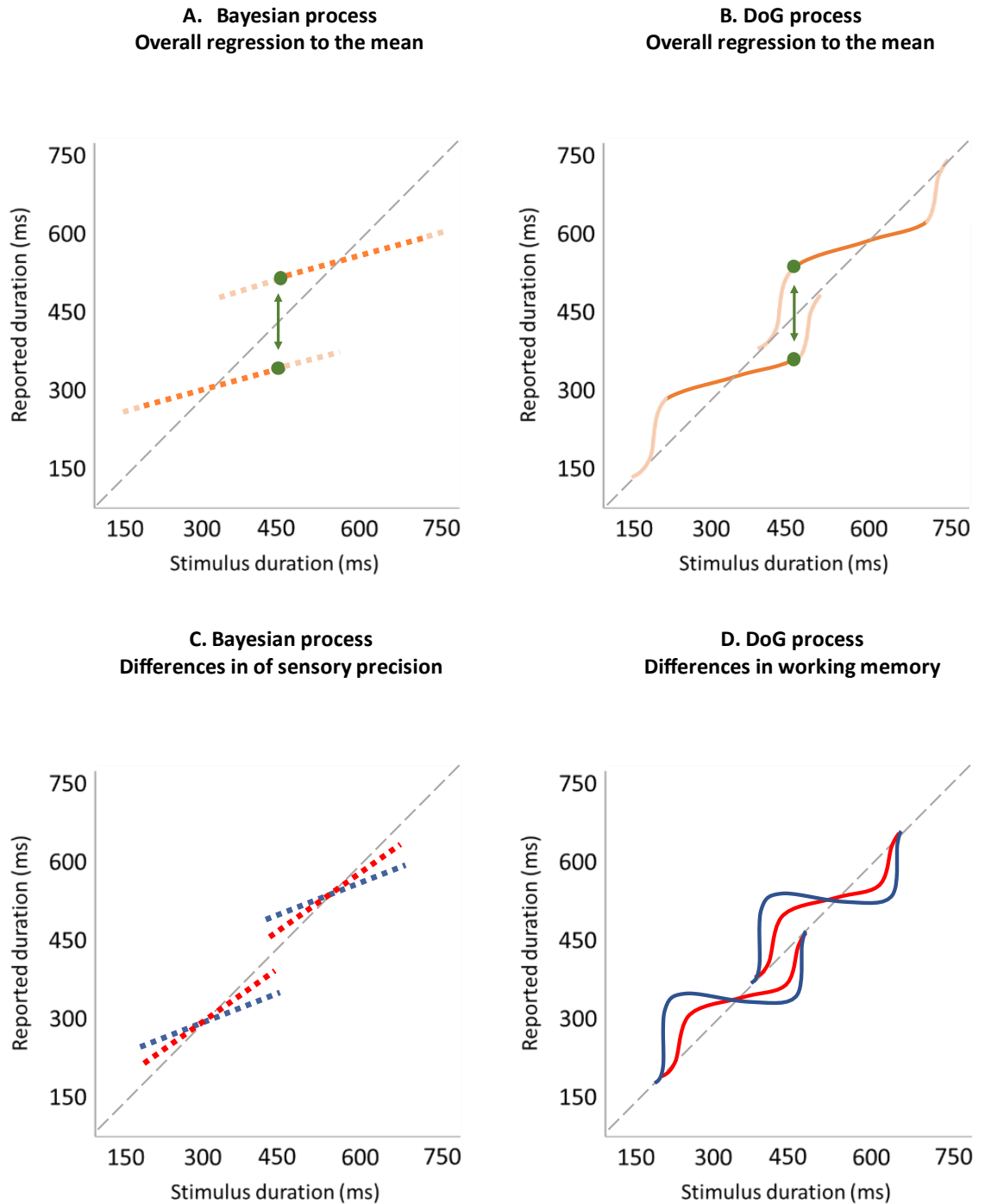


Figure 8.2. Perceptual biases explained by a Bayesian model (left) or DoG, (right). The range effect described by studies on regression to the mean, which is also predicted by Bayesian ideal-observer models (A), can also be described by the DoG curve (B). In this illustration, the predictions of the two models diverge for stimuli outside of the presented range (light orange). The green dots signify the responses to the overlapping stimulus in the two ranges. While Bayesian models predict that differences in the magnitude of regression to the mean are related to differences in sensory precision (C), under the alternative account, differences in the magnitude of regression to the mean are related to individual differences in working memory (D).

8.2.3 Generalizability and data quality issues in duration reproduction research: task performance

Experiments aiming to investigate what factors influence participants' perception of duration, first need to provide evidence that participants are able to estimate the duration of stimuli with a certain level of accuracy. If there is no sufficient correlation between presented and reported time, then we would not be able to draw any meaningful conclusions regarding the mechanism behind time perception because the absence of a correlation between stimulus duration and participants' responses would entail that changes in participants' duration responses are not related to changes in stimulus duration. Researchers might disagree on what would be the minimal correlation between stimulus duration and responses that would provide sufficient evidence that participants can perform the task. In the reproduction and estimation experiments reported in Chapters 4 and 5, participants were required to reach a Spearman correlation of 0.65 and a slope of 0.5 for their data to be included in the analysis (Figure 8.3). In magnitude estimation tasks such as these, we would expect that the magnitude of the stimulus would be the most influential factor driving participants' responses. Even though the blind application of arbitrary cut off points for interpretation of effect sizes are rarely useful, a correlation coefficient of 0.65 sits close to the upper bound of what is conventionally considered as a moderate correlation (Overholser, & Sowinski, 2008; Mukaka, 2012; Her & Wong, 2020). In addition, in the experiment detailed in Chapter 8, participants were required to reach a Pearson's correlation of 0.85 in the duration estimation training task before starting the experiment, and all tested participants reached the threshold successfully. The experiment in Chapter 6 was conducted before the experiments in Chapters 4 and 5. Because we were aware that duration reproduction tasks are harder than duration estimation tasks, we lowered the required performance thresholds from a correlation of 0.85 to a correlation of 0.65. We, therefore, believe that our exclusion criteria were liberal while at the same time, ensuring there could be no doubt that the participants were capable of estimating the presented stimuli.

Nonetheless, there are several possibilities to improve the way performance criteria are implemented in future duration reproduction studies in order to ensure that a greater proportion of participants are included in the analysis. In all experiments, we have employed "flat" exclusion criteria – if after a certain number of blocks participants are below the threshold, they are excluded from the experiment. This approach ignores individual differences in learning rates. Studies in other domains have reported individual differences in participants' use of feedback in perceptual tasks (Santesso et al., 2008; Withagen & van

Wermeskerken, 2009; Arbel & Wu, 2016; Muller-Gass, Duncan, Tavakoli, & Campbell, 2019). Evidence that differences in perceptual learning could have affected performance in our experiments comes from Chapter 4, where 12 participants did not reach the necessary performance thresholds after completing two blocks of 600 trials each. Eight of these participants showed no improvement in performance from one block to the next, suggesting that further training would be unlikely to lead to better performance. In contrast, the other four participants showed marked improvement between the first and the second block, suggesting that with further training, they would have been able to learn the task. Therefore, an alternative strategy to estimate performance that would reduce the exclusion rate would be to keep testing participants even if they are below the threshold as long as they show improvement from block to block and exclude participants only if their performance plateaus below the required performance threshold. In addition, one issue with using the correlation and slope between presented and reproduced time as performance metrics is that a low correlation or slope might reflect either random responses or complete regression to the mean, and neither the slope nor the correlation can differentiate between these two cases. Therefore, a complimentary strategy to reduce the exclusion rate while still ensuring high-quality data would be to decrease the required correlation but to add another performance metric quantifying the variability of participants' responses to each stimulus. If responses to a given duration are clustered close across trials this would indicate genuine performance even if the correlation is lower. If responses vary widely, this would indicate participants are giving responses at random.

Our results show that a very small proportion of people in the general population were able to reproduce durations, casting doubt over the reliability of the effects reported by previous studies using duration reproduction tasks to study time perception. Across the duration reproduction experiments reported in Chapters 4 and 5, we found that only 35% and 46% of participants, respectively, fulfilled the performance criteria of each study. Past studies examining perceptual biases in duration estimation do not report any exclusion criteria based on task performance, and where available, results suggest that at least a part of the participants included in the analyses were not capable of doing the task (Jazayeri & Shadlen, 2010; Acerbi et al. 2012, Cicchini et al., 2012; Aagten-Murphy et al., 2014; Murai & Yotsumoto, 2016; Roach et al., 2017; Barne et al., 2018). In some studies, the researchers reported the slope of the best fit line between presented and reported duration (Murai & Yotsumoto, 2016) or the regression index between presented and reported time ($1 - \text{slope of the best fit line}$; Cicchini et al., 2012; Aagten-Murphy et al., 2014; Barne et al., 2018), allowing the reader to

assess how accurately participants were estimating the presented stimuli. All these studies included participants with a slope below 0.4. For example, in Cicchini et al. (2012) 3 of the 6 participants with no musical experience had a slope of ~ 0.3 , and the average slope of all 6 participants was 0.48. This is problematic because it means that the reported effects are biased by the results of participants who were not able to estimate time, and therefore limits the knowledge we can obtain regarding the mechanisms behind time perception from these studies. It is also important to note that most studies do not report including any training (Cicchini et al., 2012; Aagten-Murphy et al., 2014; Roach et al., 2017; Barne et al., 2018), which means that their results are further confounded by learning effects (see Acerbi et al., 2012) and makes it increasingly unlikely that participants from the general population would have been able to perform to a satisfactory degree. The discrepancy between our results and the results of these studies raises further questions in the light of the replication crisis in psychology (Pashler & Wagenmakers, 2012; Open Science Collaboration, 2015) and the increased drive towards open science (Nosek et al., 2015; Frankenhuis & Nettle, 2018; Spellman, Gilbert & Corker, 2018). Without pre-registration of analyses and exclusion criteria, we cannot be certain how participants were recruited and selected to take part in previous studies, and we are not able to assess the suitability of the collected data to answer the research questions posed by a given study.

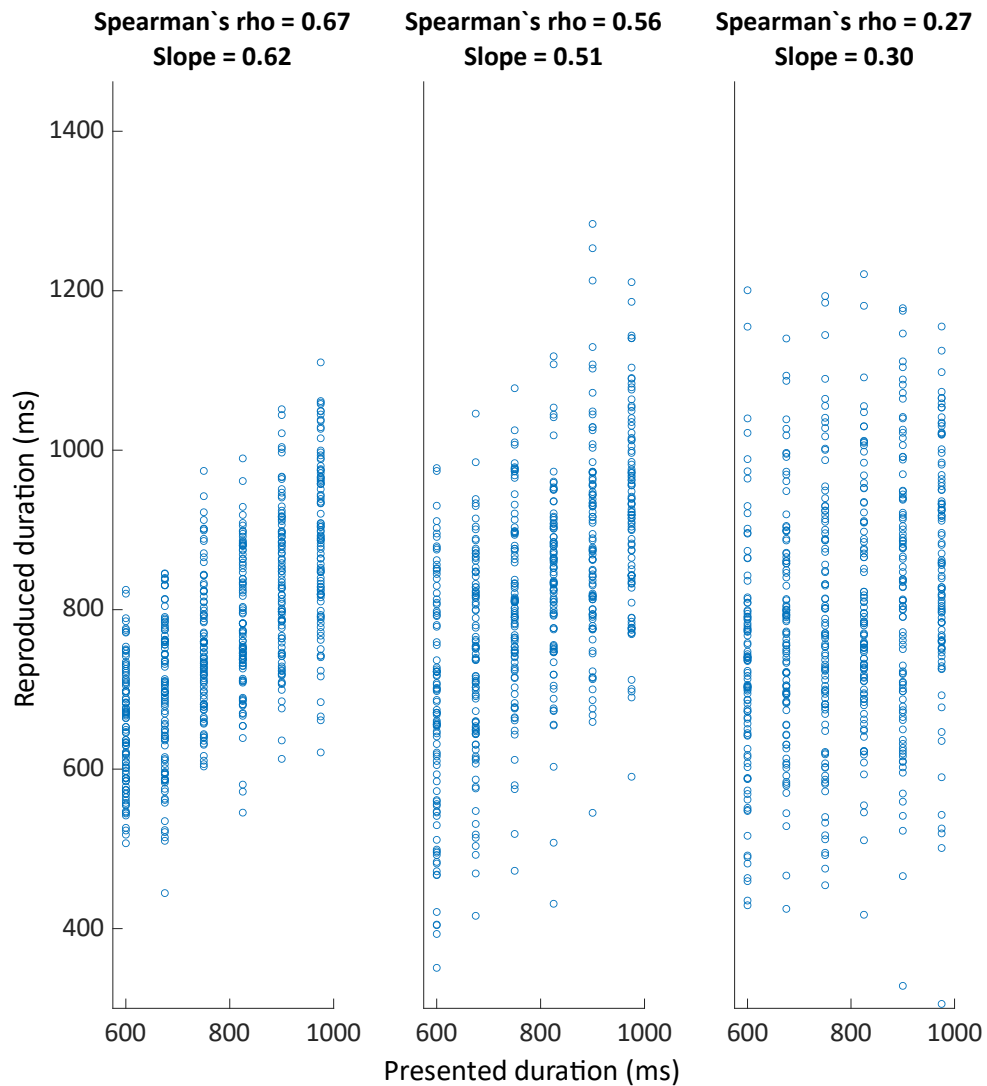


Figure 8.3. Examples of different performance in the experiment reported in Chapter 4. Each plot is based on the data from one block from the uniform condition from Chapter 4 (~570 trials), and these are three different participants. The blue dots show participants' individual responses on each trial. On the left plot, we can see that a correlation of 0.67 and slope of 0.62 (higher than our thresholds) are achievable with considerable overlap of responses across the presented stimuli (even between the shortest and the longest stimuli) – participants did not have to give completely accurate responses to reach our thresholds. Nonetheless, as the duration of the presented stimulus increased, it is clearly visible that the duration of the provided responses increased as well. As the correlation and/or slope decrease below 0.65 and 0.5 respectively (middle and right), responses become more varied and the relationship between presented and reported time becomes less clear.

While the failure of past studies to properly control for participants' ability affects the reliability of their results, the fact that more than half of the people we tested were not capable of accurately reproducing durations suggests that the findings of properly controlled reproduction studies do not generalize. However, we believe that the lack of generalizability of our findings does not stem only from the conservative exclusion criteria that we have used, but also from limitations of duration reproduction as a method to study time perception. While 46% of participants fulfilled the performance criteria required in the reproduction task in Chapter 5, 68% of participants managed to reach the same performance threshold in the duration estimation task. As we used the same stimuli in the reproduction and estimation tasks in Chapter 5 (the durations of the stimuli differed only because of difference in monitor refresh rates), the differences in performance likely reflect differences in non-timing related components of the tasks and indicate that the performance criteria alone cannot explain the high exclusion rate in the duration reproduction experiment. Further support for this idea comes from the finding that participants in the duration estimation study in Chapter 8 managed to reach a much higher performance threshold – a Pearson correlation of 0.85, although the sample in this study was much smaller. In a duration estimation task, participants need to estimate the duration of an event. In a duration reproduction task, participants need to estimate the duration of an event, and then produce a motor command with the same duration. Therefore, in duration reproduction tasks it is not clear whether we are studying participants' ability to estimate time or produce accurate motor commands, which was the main motivation for conducting the duration estimation experiment in Chapter 5. It is possible that some of the participants who failed in the reproduction tasks were capable of estimating the presented durations but failed to map their internal estimates to the appropriate motor response. In addition, if there is a discrepancy between estimated and produced duration, the feedback in reproduction tasks becomes harder to incorporate. The participant would not know whether to apply the feedback to their internal representation of the duration or the duration of the motor command. To summarize, the findings across our studies suggest that duration reproduction tasks are confounded by variables unrelated to the timing component of the task, and suffer from considerable generalizability issues. Our results show that only a small proportion of the population are likely to be capable of performing these tasks to a satisfactory degree. Therefore, albeit ensuring high data quality, the stringent exclusion criteria that we have used reduce the generalizability of our findings, which highlights the limitations of duration reproduction tasks as a method of studying time perception in the general population. If research strives to uncover the general mechanisms underlying time perception, future studies should move away from duration reproduction tasks and instead implement

tasks that fulfil some minimum generalizability criteria to re-evaluate the relevance of past findings based on potentially flawed methodology.

Finally, one might consider that our performance criteria could have led to the inclusion of only participants who perform the tasks in the experiments in Chapter 5 at ceiling, resulting in little performance variability across participants. For example, in Experiment 2, low variability in regression to the mean across participants could pose an issue for computing the correlation between bias and sensory precision. As discussed above, the metrics we used to judge performance in the duration estimation and reproduction tasks conflate poor performance and strong regression to the mean – a low correlation between presented and reported duration can mean either random responding or complete assimilation of the prior. Excluding participants who show complete regression to the mean would be an issue when computing the correlation between bias and sensory precision, because the correlation coefficient is sensitive to the range of available values. If the parameter values used to compute a correlation are restricted to a range of the parameter values available in the population, the correlation coefficient shrinks. In addition, in Experiment 1, low variability in discrimination performance could pose an issue for conducting a median split to classify participants into low and high sensory precision groups. If the participants included in the experiment have very similar 2IFC thresholds, there will only be a marginal difference between participants across the two groups. Therefore, our results would show no effect of sensory precision on regression to the mean even if such an effect exists in the population. Nonetheless, we were able to show that there was variability in regression to the mean and discrimination performance across both experiments in Chapter 5 (for the distribution of bias and sensory precision in Experiment 1 see Appendix E, Figure E1 and Appendix F, Figure F2 and F3; for the distribution of bias and sensory precision in Experiment 2 see Appendix E, Figure E2 and Appendix H). Furthermore, even if there is only little variability in participants' performance across the tasks, the experiments aimed to investigate whether Bayesian ideal observer models can predict the effect of sensory precision on regression to the mean. In other words, we were not interested in the raw magnitude of the effect of precision on bias, but whether the magnitude of the effect in the behavioural data corresponded to the magnitude of the effect predicted by the models. Importantly, the models used participants' 2IFC data as input, therefore incorporating the potentially limited range in participant performance. If the models predicted no effect of precision on bias and we found no effect of precision on bias in the behavioural data, then the behavioural data would have supported the predictions of the model. However, in the results of both experiments in Chapter 5, the relationship between precision and bias in the

behavioural data did not support the relationship predicted by the Bayesian ideal observer models. Even given the potentially restricted range of sensory precision values in the data, the models in both experiments predicted a very strong relationship between sensory precision and regression to the mean. As discussed in Chapter 5 (p. 108), in Experiment 2 we used a model with an updating prior because our original model, which assumed a fixed prior, predicted a nearly perfect correlation between bias and precision. Therefore, while more lenient exclusion criteria would have resulted in greater variability in performance across participants, this would have had little effect on our inferential conclusions as the variability in participants' performance is accounted for in the predictions of the models.

8.2.4 Data quality issues in duration reproduction research: equipment timing accuracy

A final issue in the duration reproduction literature is that most researchers do not specify the latency of the equipment used to record participants' responses, so it is impossible to determine the amount of measurement noise corrupting the data (see Acerbi et al. 2012 for a notable exception). In the duration reproduction study reported in Chapter 5, we used a keyboard advertised as a high-speed gaming keyboard to record participants reproduced durations of the presented stimuli. In addition, we used the KbQueue family of functions in Psychtoolbox as they have been reported to detect the timing of keypresses more accurately than the standard Kb functions (KbQueueCreate, 2020). Our initial plan was to use the same keyboard in the reproduction experiment detailed in Chapter 4⁴. However, as part of the stage 1 review process of the registered report, a reviewer drew our attention to possible latency issues with recording responses on keyboards. Therefore, we recorded the latency between the time at which a participant presses the space key on the keyboard and the time at which the computer records that the key was pressed. We attached a pressure sensor to the space bar of the keyboard and connected the pressure sensor to an external audio card with 44100Hz refresh rate (M-Audio Delta 66). That way, we could compare the true time of button presses, recorded by the audio card, with the timing of the button press recorded by the computer. One participant (the author) generated 1000 button presses, and we found that, on average, the computer registered a keypress 16.24ms (SD = 3.82) later than the pressure sensor. In addition, there was considerable variability in the recorded latencies, with the highest recorded latency reaching 31.34ms (Figure 8.4A). The problem is not so much in the latency itself – if the keypresses recorded by the computer are always late by a consistent,

⁴ Data collection for Experiment 1 in Chapter 5 preceded data collection for the experiments in Chapter 4.

fixed amount, this latency can be simply subtracted from participants' responses. If the latency varies from trial to trial, however, then it becomes difficult to account for. Based on these results, we can be fairly certain that participants' duration reproductions were corrupted by variable measurement noise which could have obscured the relationship between sensory precision and regression to the mean. Nonetheless, Experiment 2, where we used direct estimation instead of reproduction, suggests that measurement noise alone is not enough to explain the null result in the reproduction study.

Based on the results of the keyboard latency test above, in Chapter 4 we recorded participants responses using a custom-built button box connected to a professional-grade audio-card (M-Audio Delta 66; 44 100Hz refresh rate), which allowed us to record the duration of participants' reproductions with high accuracy. Nonetheless, we had to ensure that the duration of the stimuli was presented accurately. This was not a trivial issue as in that experiment, the start of the interval was defined by a button press and the end of the interval by the presentation of a stimulus on the screen. Therefore, we conducted a timing calibration experiment prior to data collection to measure how accurately the computer presented the desired interval between a button press on the button box and the presentation of a stimulus on the monitor. The button box was connected to two computers equipped with professional-grade audio-cards – the computer on which participants would subsequently complete the experiment (the presentation computer), and a second computer which was used to measure the latency between a button press and the presentation of a stimulus (the measurement computer). The timing of stimulus presentation on the presentation computer was measured using a photodiode which was connected to the audio-card of the measurement computer. The calibration experiment was carried out in a dark room. On every trial, one participant (the author) generated a press on the button box. After some pre-specified time (equal to the intervals used in the experiment) a white circle was presented on the screen of the presentation computer, which was otherwise black. We then measured the difference in the desired and achieved interval between the button press and the visual stimulus measured by the measurement computer. The data, based on 1081 button presses⁵, shows that the average difference between desired and achieved duration was 0.82ms (SD = 2.01), and latencies were consistent across the different stimulus durations (Figure 8.4B). In addition, we found that the difference between desired and achieved duration was highest after the button

⁵ The number of button presses generated for each stimulus was as follows: 600ms – 175 button presses; 675ms – 171 button presses; 750ms – 183 button presses; 825ms – 131 button presses; 900ms – 175 button presses; 975ms – 246 button presses.

box had been inactive for a prolonged amount of time (for example, at the beginning of a testing block), so in the real experiment, the first trial in each block was discarded and participants were instructed to press the button a couple of times before restarting the experiment after a break.

Unfortunately, measuring the timing accuracy of the equipment used in experiments appears to not be a common practice in psychology research, which has led to repeated calls for researchers to exercise greater control and transparency (Plant & Quinlan, 2013; Plant, 2016). This is particularly problematic given that previous studies have documented great variability in keyboard latencies not only within but also between devices. For example, a study testing 11 keyboards showed that average latencies across devices ranged from 4.24ms to 49.04ms (Figure 8.4C, Wimmer, Schmid, & Bockes, 2019), and similar findings have been reported elsewhere (Luu, 2017; Bockes, Wimmer, & Schmid, 2018). Altogether these results show that keyboards do not generally provide the sufficient accuracy needed to record responses where the timing of button presses is of crucial importance. However, most studies using duration reproduction or ready-set-go tasks to study perceptual biases in time perception use keyboards to collect participant's responses and do not provide any information regarding their latencies (Jazayeri & Shadlen, 2010; Cicchini et al. 2012; Aagten-Murphy et al., 2014; Roach et al., 2017). Others do not even specify what equipment was used to collect responses (Murai & Yotsumoto, 2016; Barne et al., 2018), although it is probably safe to assume that the researchers would have indicated if they used specialized equipment or conducted any calibration tests. This issue should not be taken lightly, as the failure to validate timing accuracies prior to data collection can invalidate a study's results (e.g., Crosse & Lalor, 2014). We do not believe that the duration reproduction studies that have not reported equipment latencies have done so in bad faith, as we were originally not aware of how severe the latency failure of modern keyboards was either. However, it is the responsibility of the researcher to measure and report the timing accuracy of the equipment used and not proceed with the planned experiment unless stimuli can be presented, and responses recorded with sufficient accuracy. This is especially important for research pertaining to time perception. It should become standard practice for researchers to record and report the timing accuracy of equipment in scientific articles, and for reviewers to demand this information be included when missing. If researchers do not exercise stringent controls, it becomes easy to see why psychology overall has a poor reputation amongst other disciplines and is not considered a proper science by many (Berezow, 2012).

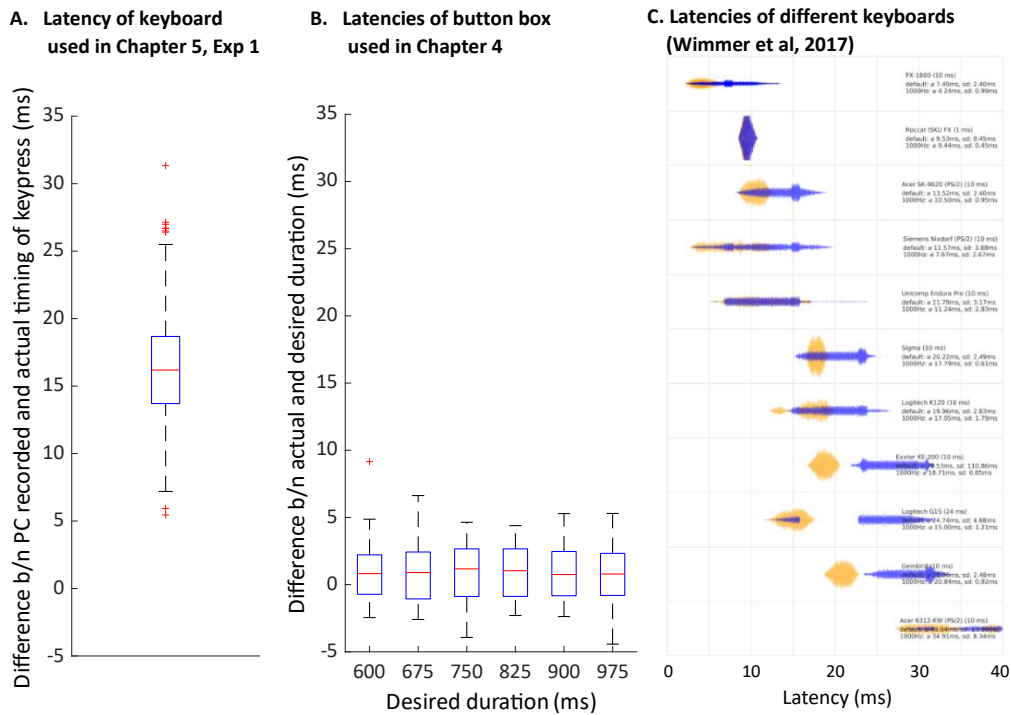


Figure 8.4. Distribution of device latencies. (A) shows the results of our timing experiment with the keyboard used to reproduce durations in Chapter 5. The boxplot shows the distribution of differences in the time of keypresses recorded by the audio card and the time of the keypresses recorded by the PC. Negative differences mean that the time recorded by the computer was later than the time recorded by the audio card. The median difference was 16.19ms, and in 50% of the recorded keypresses, the difference was between 13.70 and 18.68ms. The red crosses represent outliers. (B) shows the difference between desired and achieved interval duration for each stimulus using the button box from Chapter 5. Across all stimuli, the median difference was 0.90ms, and 50% of the differences were between -0.85 and 2.47ms. The outlier (marked in red) occurred at the start of a testing block, likely because the button box was inactive for a long period of time. Therefore, to reduce the occurrence of such errors, we discarded the first trial in each block. (C) shows the latencies distributions of 11 different keyboards recorded by Wimmer et al. (2017). Blue shows latencies recorded using default USB polling rates, and orange shows latencies using 1000Hz polling rate (roughly, polling rate refers to how often the computer checks the USB device for new data). Figure adapted from Wimmer et al. (2017).

8.3 Future directions

In the discussion above, I identified some issues that require further research – the interaction between attractive and repulsive biases in perceptual reports, and the role of working memory in global perceptual biases. In addition to these issues, the findings of the research presented here raised a number of additional questions regarding the neural correlates of perceptual biases in magnitude estimation that future research should address. First, as regression to the mean is a general phenomenon documented across multiple sensory domains and tasks, it would be worth examining if the EEG correlate of regression to the mean we reported in Chapter 5 also generalizes across modalities. Furthermore, if brain activity reflects the range of

the underlying stimulus distribution, it would also be intriguing to investigate whether and how the effect of distributional variance on behaviour is reflected in the brain.

8.3.1 Do the neuronal correlates of perceptual biases in duration estimation generalise to magnitude estimation in other domains?

In Chapter 5, we provide evidence of EEG correlates for contextual effects in duration estimation. One potential direction for future research to extend on these findings will be to investigate if the neural signatures of these contextual effects generalise across modalities. Walsh' (2003) "A Theory Of Magnitude" (ATOM) suggests that the brain encodes magnitude as an abstract, unitless feature, independent of the actual dimension (for example size or duration) being judged. Evidence that magnitude estimation across dimensions is based on a common mechanism is provided by studies reporting that temporal judgements are affected by differences in magnitude-based characteristics of the stimulus that are not relevant to the task – for example, bigger stimuli are judged as longer than smaller stimuli presented for the same duration (Xuan, Zhang, He, & Chen, 2007). Similar effects on temporal judgements have been reported for the number of presented items, luminosity, and numerical value when the stimuli are Arabic digits (Xuan et al., 2007; Oliveri et al., 2008). Neuroimaging studies provide further evidence for overlapping activation in parietal and frontal brain areas across different magnitude estimation tasks (Buelti & Wash, 2009; Koch, Oliveri, & Caltagirone, 2009; Sokolowski, Fias, Ononye, & Ansari, 2017). Studies have reported overlap in activation across parietal and frontal areas for judgements on duration and numerosity of stimuli (Hayashi et al., 2013), judgements on the physical and numerical magnitude of digits (Kaufmann et al., 2005), as well as judgements on duration, numerosity and length (Skagerlund, Karlsson, Träff, 2016). Nonetheless, multiple studies report findings challenging the idea that magnitude estimation is based on an overlapping sensory processing mechanism. For example, behavioural studies have found that interference between relevant and irrelevant stimulus dimensions in magnitude estimation tasks is not always bidirectional – while the numerosity of the presented items has been found to interfere with judgements of duration and length, the duration or length of the stimulus does not interfere with judgements of numerosity (Dormal, Seron, & Pesenti, 2006; Dormal, & Pesenti, 2007). In addition, at least some of the positive results in the literature (e.g. Xuan et al., 2007; Oliveri et al., 2008) can be the result of decision rather than perceptual bias – changes in a task-irrelevant dimension (e.g. size) could bias decisions regarding another stimulus dimension (e.g. duration) without changing how the task-relevant dimension of the stimulus is actually perceived (Yates, Loetscher, & Nicholls, 2012).

Based on the conflicting findings in the literature, further research is needed to investigate whether a common magnitude perception mechanism exists across domains. We provide evidence that contextual effects in duration estimation, whether a stimulus belongs to the distribution of long or short intervals, can be decoded using EEG (also see Damsma et al., 2020 for further supporting evidence). Behavioural studies to date have reported that regression to the mean is present across many perceptual domains such as duration, size, distance (e.g., Jazayeri & Shadlen, 2010; Hollingworth, 1910; Petzschner & Glasauer, 2011) but no studies have yet directly investigated the relationship between these effects. Therefore, one question emerging from these findings is whether behavioural and neuroimaging contextual effects in one domain (e.g., duration) would predict contextual effects in another domain (e.g., numerosity).

8.3.2 What are the neural correlates of the variance effect on perceptual biases in behaviour?

Another potential area for future research based on our findings would be to investigate what neural correlates reflect the effect of variance on perceptual biases. Previous studies have shown that the amplitude of the MMN elicited by a deviant stimulus depends on the variance and range of the underlying stimulus distribution (Garrido et al., 2013, 2016). In Garrido et al.'s (2013) study, participants were presented with auditory stimuli whose pitch was drawn from a normal distribution with either small variance and narrow range, or large variance and wide range. The researchers designed a paradigm in which the deviant and standard stimuli are defined probabilistically, rather than based on the frequency of their presentation. In most of the trials, for example, 80% of trials, participants are presented with stimuli drawn from a normal distribution with a given mean and standard deviation (but excluding the mean). The remaining 20% of the trials contain the standard (10%) and deviant (10%) stimuli. The standard stimulus is defined as the mean of the distribution (meaning that the standard is statistically very likely to appear), while the deviant is drawn from the tail of the distribution (meaning that the deviant is statistically very unlikely to appear). An important feature of this paradigm which contrasts it from standard oddball tasks used in the field (Näätänen, Pakarinen, Rinne, & Takegata, 2004) is that the standard and deviant stimuli are presented for the same number of trials – what defines which stimulus is perceived a standard and which as deviant depends on the probabilistic structure imposed by the filler trials. Using this task, Garrido et al. (2013) found that the amplitude of the MMN in response to the unexpected stimulus was greater when the stimuli were drawn from the distribution with small variance and narrow range rather than a distribution with large variance and wide range. However, as the researchers

manipulated the variance and range of the distributions concurrently is it difficult to draw definite conclusions about the specific effects of variance on the amplitude of the MMN. Here we showed that when keeping range constant, behavioural responses are sensitive to change in variance alone (Chapter 4). Therefore, a natural next step in our research would be to investigate if MMN would also reflect unconfounded changes in the variance of stimulus distributions.

8.4 Conclusions

The main aim of the research presented in this thesis was to investigate if duration perception follows the principles of Bayesian perceptual inference. The reported results show that behavioural responses are sensitive to the variance of the underlying stimulus distribution, providing direct evidence for a fundamental prediction of Bayesian perceptual theories. Nonetheless, our findings also highlight the need for future research to focus on developing a comprehensive framework designed to account for a wider range of opposing perceptual biases. There is also a gap in the literature concerning how factors not usually considered in Bayesian perceptual models, for example working memory, could influence attractive and repulsive biases. This work also draws attention to the importance of exercising greater transparency and data quality control in empirical research as we can learn little about perception from corrupted data and tasks that do not generalize.

References

- Aagten-Murphy, D., Cappagli, G., & Burr, D. (2014). Musical training generalises across modalities and reveals efficient and adaptive mechanisms for reproducing temporal intervals. *Acta psychologica*, 147, 25-33.
- Abel, S. M. (1972). Duration discrimination of noise and tone bursts. *The Journal of the Acoustical Society of America*, 51(4B), 1219-1223.
- Ablin, P., Cardoso, J. F., & Gramfort, A. (2018). Faster independent component analysis by preconditioning with Hessian approximations. *IEEE Transactions on Signal Processing*, 66(15), 4040-4049.
- Acerbi, L., Vijayakumar, S., & Wolpert, D. M. (2014). On the origins of suboptimality in human probabilistic inference. *PLoS computational biology*, 10(6), e1003661.
- Acerbi, L., Wolpert, D. M., & Vijayakumar, S. (2012). Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS computational biology*, 8(11), e1002771.
- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, 14(3), 257-262.
- Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science*, 21(4), 560-567.
- Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, Perception, & Psychophysics*, 74(5), 810-815.
- Allan, L. G. (1992). The internal clock revisited. In *Time, action and cognition* (pp. 191-202). Springer, Dordrecht.
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological science*, 19(4), 392-398.
- Anderson, N. H. (1967). Application of a weighted average model to a psychophysical averaging task. *Psychonomic Science*, 8(6), 227-228.
- Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2011). Visual and vestibular cue integration for heading perception in extrastriate visual cortex. *The Journal of physiology*, 589(4), 825-833.
- Arbel, Y., & Wu, H. (2016). A neurophysiological examination of quality of learning in a feedback-based learning task. *Neuropsychologia*, 93, 13-20.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological science*, 12(2), 157-162.
- Atienza, M., Cantero, J. L., & Gómez, C. M. (1997). The mismatch negativity component reveals the sensory memory during REM sleep in humans. *Neuroscience letters*, 237(1), 21-24.

- Atienza, M., & Cantero, J. L. (2001). Complex sound processing during human REM sleep by recovering information from long-term memory as revealed by the mismatch negativity (MMN). *Brain research*, 901(1-2), 151-160.
- Avant, L. L., Lyman, P. J., & Antes, J. R. (1975). Effects of stimulus familiarity upon judged visual duration. *Perception & Psychophysics*, 17(3), 253-262.
- Balci, F. (2014). Interval timing, dopamine, and motivation. *Timing & Time Perception*, 2(3), 379-410.
- Bannert, M. M., & Bartels, A. (2013). Decoding the yellow of a gray banana. *Current Biology*, 23(22), 2268-2272.
- Barne, L. C., Sato, J. R., de Camargo, R. Y., Claessens, P. M., Caetano, M. S., & Cravo, A. M. (2018). A common representation of time across visual and auditory modalities. *Neuropsychologia*, 119, 223-232.
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Josa a*, 20(7), 1391-1397.
- Baykova, R. & Roseboom, W. (2019). Effect of sensory precision on behavioural and neuroimaging perceptual biases in duration estimation. *CCN 2019 Conference proceedings*. Available at: <https://doi.org/10.32470/CCN.2019.1280-0>
- Baykova, R., Buckley, C., Seth, A., & Roseboom, W. (in-principle acceptance). Influence of prior distribution variance on time perception (Registered Report Stage 1). *Royal Society Open Science*. Available at: <https://osf.io/75p42/>
- Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., ... & Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, 60(6), 1142-1152.
- Berezow, A. (2012). Why psychology isn't a science. *Los Angeles Times*.
- Berniker, M., Voss, M., & Kording, K. (2010). Learning priors for Bayesian computations in the nervous system. *PloS one*, 5(9), e12686.
- Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific reports*, 7(1), 1-13.
- Block, R. A. (1974). Memory and the experience of duration in retrospect. *Memory & Cognition*, 2(1), 153-160.
- Bockes, F., Wimmer, R., & Schmid, A. (2018, April). LagBox--Measuring the Latency of USB-Connected Input Devices. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-6).
- Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological bulletin*, 138(3), 389.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10, 433-436.
- Bristow, D., Frith, C., & Rees, G. (2005). Two distinct neural effects of blinking on human visual processing. *Neuroimage*, 27(1), 136-145.
- Brown, S. W., & Stubbs, D. A. (1988). The psychophysics of retrospective and prospective timing. *Perception*, 17(3), 297-310.

- Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1525), 1831-1840.
- Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature reviews neuroscience*, 6(10), 755-765.
- Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *IEEE transactions on medical imaging*, 18(1), 32-42.
- Cantlon, J. F., Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on "Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures". *science*, 323(5910), 38-38.
- Casini, L., & Vidal, F. (2011). The SMAs: neural substrate of the temporal accumulator?. *Frontiers in integrative neuroscience*, 5, 35.
- Cavanagh, J. F., Masters, S. E., Bath, K., & Frank, M. J. (2014). Conflict acts as an implicit cost in reinforcement learning. *Nature communications*, 5(1), 1-10.
- Chang, C. J., & Jazayeri, M. (2018). Integration of speed and time for estimating time to contact. *Proceedings of the National Academy of Sciences*, 115(12), 2879-2887.
- Chang, A. Y. C., Seth, A. K., & Roseboom, W. (2017). Neurophysiological signatures of duration and rhythm prediction across sensory modalities. *BioRxiv*, 183954.
- Chen, Y., Huang, X., Luo, Y., Peng, C., & Liu, C. (2010). Differences in the neural basis of automatic auditory and visual time perception: ERP evidence from an across-modal delayed response oddball task. *Brain research*, 1325, 100-111.
- Chong, S. C., Joo, S. J., Emmanouil, T. A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception & Psychophysics*, 70(7), 1327-1334.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision research*, 43(4), 393-404.
- Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision research*, 45(7), 891-900.
- Church, R. M., & Meck, W. H. (2003). A concise introduction to scalar timing theory. *Functional and neural mechanisms of interval timing*, 3-22.
- Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences*, 111(21), 7867-7872.
- Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *Journal of Neuroscience*, 32(3), 1056-1060.
- Cicchini, G. M., Mikellidou, K., & Burr, D. C. (2018). The functional role of serial dependence. *Proceedings of the Royal Society B*, 285(1890), 20181722.

- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, 36(3), 181-204.
- Clifford, C. W., & Wenderoth, P. (1999). Adaptation to temporal modulation can enhance differential speed sensitivity. *Vision research*, 39(26), 4324-4331.
- Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta psychologica*, 138(2), 289-301.
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613-617.
- Coull, J. T., Cheng, R. K., & Meck, W. H. (2011). Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology*, 36(1), 3-25.
- Czoschke, S., Fischer, C., Beitner, J., Kaiser, J., & Bledowski, C. (2019). Two types of serial dependence in visual working memory. *British Journal of Psychology*, 110(2), 256-267.
- Crosse, M. J., & Lalor, E. C. (2014). The cortical representation of the speech envelope is earlier for audiovisual speech than audio speech. *Journal of neurophysiology*, 111(7), 1400-1408.
- Curtis, D. W., & Mullin, L. C. (1975). Judgments of average magnitude: Analyses in terms of the functional measurement and two-stage models. *Perception & Psychophysics*, 18(4), 299-308.
- Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision research*, 37(22), 3181-3192.
- Damsma, A., Schlichting, N., & van Rijn, H. (2020). Temporal context actively shapes EEG signatures of time perception. *bioRxiv*.
- Dang, L. C., Samanez-Larkin, G. R., Castrellon, J. J., Perkins, S. F., Cowan, R. L., Newhouse, P. A., & Zald, D. H. (2017). Spontaneous eye blink rate (EBR) is uncorrelated with dopamine D2 receptor availability and unmodulated by dopamine agonism in healthy adults. *Eneuro*, 4(5).
- Darlington, T. R., Beck, J. M., & Lisberger, S. G. (2018). Neural implementation of Bayesian inference in a sensorimotor behavior. *Nature neuroscience*, 21(10), 1442.
- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9-21.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in psychology*, 5, 781.
- Donchin, E., Gerbrandt, L. A., Leifer, L., & Tucker, L. (1972). Is the contingent negative variation contingent on a motor response?. *Psychophysiology*, 9(2), 178-188.
- Dormal, V., & Pesenti, M. (2007). Numerosity-length interference: A Stroop experiment. *Experimental Psychology*, 54(4), 289-297.

- Dormal, V., Seron, X., & Pesenti, M. (2006). Numerosity-duration interference: A Stroop experiment. *Acta psychologica*, 121(2), 109-124.
- Droit-Volet, S. (2003). Alerting attention and time perception in children. *Journal of Experimental Child Psychology*, 85(4), 372-384.
- Droit-Volet, S., Clément, A., & Fayol, M. (2003). Time and number discrimination in a bisection task with a sequence of stimuli: A developmental approach. *Journal of experimental child psychology*, 84(1), 63-76.
- Droit-Volet, S., & Wearden, J. (2002). Speeding up an internal clock in children? Effects of visual flicker on subjective duration. *The Quarterly Journal of Experimental Psychology: Section B*, 55(3), 193-211.
- Dormal, V., Seron, X., & Pesenti, M. (2006). Numerosity-duration interference: A Stroop experiment. *Acta psychologica*, 121(2), 109-124.
- Durstewitz, D. (2003). Self-organizing neural integrator predicts interval times through climbing activity. *Journal of Neuroscience*, 23(12), 5342-5353.
- De Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, 108(32), 13341-13346.
- de Gardelle, V., & Mamassian, P. (2015). Weighting mean and variability during confidence judgments. *PloS one*, 10(3), e0120870.
- Eckner, J. T., Kutcher, J. S., & Richardson, J. K. (2010). Pilot evaluation of a novel clinical test of reaction time in National Collegiate Athletic Association Division I football players. *Journal of athletic training*, 45(4), 327-332.
- Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American statistical Association*, 82(397), 171-185.
- Egger, S. W., Le, N. M., & Jazayeri, M. (2020). A neural circuit model for human sensorimotor timing. *Nature communications*, 11(1), 1-14.
- Elliott, R. (1968). Simple visual and simple auditory reaction time: A comparison. *Psychonomic Science*, 10(10), 335-336.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429.
- Fahrenfort, J. J., Van Driel, J., Van Gaal, S., & Olivers, C. N. (2018). From ERPs to MVPA using the Amsterdam decoding and modeling toolbox (ADAM). *Frontiers in neuroscience*, 12, 368.
- Fechner, G. T. (1860). *Elemente der psychophysik* (Vol. 2). Breitkopf u. Härtel.
- Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, 63(4), 719-725.
- Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2010). Visual-vestibular cue integration for heading perception: applications of optimal cue integration theory. *European Journal of Neuroscience*, 31(10), 1721-1729.

- Fetterman, J. G., & Killeen, P. R. (1992). Time discrimination in *Columba livia* and *Homo sapiens*. *Journal of Experimental Psychology: Animal Behavior Processes*, 18(1), 80.
- Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature neuroscience*, 17(5), 738-743.
- Flynn, M., Liasis, A., Gardner, M., & Towell, T. (2017). Visual mismatch negativity to masked stimuli presented at very brief presentation rates. *Experimental brain research*, 235(2), 555-563.
- Fraisse, P. (1952). Perception of duration as organization of successive stimuli; experimental demonstration. *L'annee psychologique*, 52(1), 39-46.
- Frankenhaeuser, M. (1959). *Estimation of time: An experimental study*.
- Frankenhuis, W. E., & Nettle, D. (2018). Open science is liberating and can foster creativity. *Perspectives on Psychological Science*, 13(4), 439-447.
- Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, 11(2), 127-139.
- Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology*, 27(4), 590-595.
- Gaillard, A. W., & Näätänen, R. (1973). Slow potential changes and choice reaction time as a function of interstimulus interval. *Acta Psychologica*, 37(3), 173-186.
- Gallivan, J. P., & Chapman, C. S. (2014). Three-dimensional reach trajectories as a probe of real-time decision-making between multiple competing targets. *Frontiers in neuroscience*, 8, 215.
- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical neurophysiology*, 120(3), 453-463.
- Garrido, M. I., Sahani, M., & Dolan, R. J. (2013). Outlier responses reflect sensitivity to statistical structure in the human brain. *PLoS computational biology*, 9(3).
- Garrido, M. I., Teng, C. L. J., Taylor, J. A., Rowe, E. G., & Mattingley, J. B. (2016). Surprise responses in the human brain demonstrate statistical learning under high concurrent cognitive demand. *npj Science of Learning*, 1, 16006.
- Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. *Perception & Psychophysics*, 18(1), 1-8.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological review*, 84(3), 279.
- Glasauer, S., & Shi, Z. (2018). 150 years of research on Vierordt's law-Fechner's fault?. *bioRxiv*, 450726.
- Golan, T., Davidesco, I., Meshulam, M., Groppe, D. M., Mégevand, P., Yeagle, E. M., ... & Deouell, L. Y. (2016). Human intracranial recordings link suppressed transients rather than 'filling-in' to perceptual continuity across blinks. *Elife*, 5, e17243.
- Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 290(1038), 181-197.

- Groman, S. M., James, A. S., Seu, E., Tran, S., Clark, T. A., Harpster, S. N., ... & Elsworth, J. D. (2014). In the blink of an eye: relating positive-feedback sensitivity to striatal dopamine D2-like receptors through blink rate. *Journal of Neuroscience*, 34(43), 14443-14454.
- Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & psychophysics*, 54(3), 383-394.
- Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, 72(3), 561-582.
- Grondin, S. (2012). Violation of the scalar property for time perception between 1 and 2 seconds: evidence from interval discrimination, reproduction, and categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 880.
- Grondin, S., Gamache, P. L., Tobin, S., Bisson, N., & Hawke, L. (2008). Categorization of brief temporal intervals: An auditory processing context may impair visual performances. *Acoustical science and technology*, 29(5), 338-340.
- Grondin, S., & Killeen, P. R. (2009). Tracking time with song and count: Different Weber functions for musicians and non-musicians. *Attention, Perception, & Psychophysics*, 71(7), 1649-1654.
- Grondin, S., Ouellet, B., & Roussel, M. È. (2001). About optimal timing and stability of Weber fraction for duration discrimination. *Acoustical science and technology*, 22(5), 370-372.
- Grondin, S., Ouellet, B., & Roussel, M. E. (2004). Benefits and limits of explicit counting for discriminating temporal intervals. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 58(1), 1.
- Groppe, D. M., Urbach, T. P., & Kutas, M. (2011a). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, 48(12), 1711-1725.
- Groppe, D. M., Urbach, T. P., & Kutas, M. (2011b). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology*, 48(12), 1726-1737.
- Grossman, S., Gueta, C., Pesin, S., Malach, R., & Landau, A. N. (2019). Where does time go when you blink?. *Psychological science*, 30(6), 907-916.
- Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision*, 15(4), 16-16.
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current biology*, 17(17), R751-R753.
- Haberman, J., & Whitney, D. (2009a). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 718.
- Haberman, J., Harp, T., & Whitney, D. (2009b). Averaging facial expression over time. *Journal of vision*, 9(11), 1-1.
- Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature neuroscience*, 9(11), 1367-1368.

- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.
- Hayashi, M. J., Kanai, R., Tanabe, H. C., Yoshida, Y., Carlson, S., Walsh, V., & Sadato, N. (2013). Interaction of numerosity and time in prefrontal and parietal cortex. *Journal of Neuroscience*, 33(3), 883-893.
- He, S., & MacLeod, D. I. (2001). Orientation-selective adaptation and tilt after-effect from invisible patterns. *Nature*, 411(6836), 473-476.
- Helbig, H. B., Ernst, M. O., Ricciardi, E., Pietrini, P., Thielscher, A., Mayer, K. M., ... & Noppeney, U. (2012). The neural mechanisms of reliability weighted integration of shape information from vision and touch. *Neuroimage*, 60(2), 1063-1072.
- Heller, M. A. (1983). Haptic dominance in form perception with blurred vision. *Perception*, 12(5), 607-613.
- Helmholtz, H. V. (1860). Theorie der Luftschwingungen in Röhren mit offenen Enden. *Journal für die reine und angewandte Mathematik*, 57, 1-72.
- Henry, F. M. (1948). Discrimination of the duration of a sound. *Journal of Experimental Psychology*, 38(6), 734.
- Her, Q. L., & Wong, J. (2020). Significant correlation versus strength of correlation. *American Journal of Health-System Pharmacy*, 77(2), 73-75.
- Heron, J., Aaen-Stockdale, C., Hotchkiss, J., Roach, N. W., McGraw, P. V., & Whitaker, D. (2011). Duration channels mediate human time perception. *Proceedings of the Royal Society B: Biological Sciences*, 279(1729), 690-698.
- Hess, R. F., & Holliday, I. E. (1992). The coding of spatial position by the human visual system: effects of spatial scale and contrast. *Vision Research*, 32(6), 1085-1097.
- Ho, H. N., Chow, H. M., Tsunokake, S., & Roseboom, W. (2019). Thermal-tactile integration in object temperature perception. *IEEE transactions on haptics*.
- Hocherman, S., & Ben-Dov, G. (1979). Modality-specific effects on discrimination of short empty time intervals. *Perceptual and motor skills*, 48(3), 807-814.
- Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
- Hollingworth, H. L. (1910). The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, 7(17), 461-469.
- Howard, C. D., Li, H., Geddes, C. E., & Jin, X. (2017). Dynamic nigrostriatal dopamine biases action selection. *Neuron*, 93(6), 1436-1450.
- Hsieh, P. J., Colas, J. T., & Kanwisher, N. (2011). Pop-out without awareness: Unseen feature singletons capture attention only when top-down attention is available. *Psychological science*, 22(9), 1220-1226.
- Hsu, W. Y., Cheng, C. H., Lin, H. C., Liao, K. K., Wu, Z. A., Ho, L. T., & Lin, Y. Y. (2010). Memory-based mismatch response to changes in duration of auditory stimuli: An MEG study. *Clinical neurophysiology*, 121(10), 1744-1750.

- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1), 215-243.
- Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of vision*, 15(4), 5-5.
- Hudson, T. E., Maloney, L. T., & Landy, M. S. (2007). Movement planning with probabilistic target information. *Journal of neurophysiology*, 9, 3034-3046.
- Jack, B. N., Widmann, A., O'Shea, R. P., Schröger, E., & Roeber, U. (2017). Brain activity from stimuli that are not perceived: Visual mismatch negativity during binocular rivalry suppression. *Psychophysiology*, 54(5), 755-763.
- Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature neuroscience*, 13(8), 1020-1026.
- Jazayeri, M., & Shadlen, M. N. (2015). A neural mechanism for sensing and reproducing a time interval. *Current Biology*, 25(20), 2599-2609.
- Jongkees, B. J., & Colzato, L. S. (2016). Spontaneous eye blink rate as predictor of dopamine-related cognitive function—A review. *Neuroscience & Biobehavioral Reviews*, 71, 58-82.
- Juni, M. Z., Gureckis, T. M., & Maloney, L. T. (2012). Effective integration of serially presented stochastic cues. *Journal of vision*, 12(8), 12-12.
- Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: encoding time in neural network states. *Neuron*, 53(3), 427-438.
- Karson, C. N. (1983). Spontaneous eye-blink rates and dopaminergic systems. *Brain*, 106(3), 643-653.
- Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., ... & Ischebeck, A. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: an event-related fMRI study. *Neuroimage*, 25(3), 888-898.
- KbQueueCreate (2020). Psychtoolbox-3. . Retrieved December 21, 2020, from <http://psychtoolbox.org/docs/KbQueueCreate>
- Khayat, N., & Hochstein, S. (2018). Perceiving set mean and range: Automaticity and precision. *Journal of Vision*, 18(9), 23-23.
- Killeen, P. R., & Weiss, N. A. (1987). Optimal timing and the Weber function. *Psychological Review*, 94(4), 455.
- Kim, S., Burr, D., Cicchini, G. M., & Alais, D. (2020). Serial dependence in perception requires conscious awareness. *Current Biology*, 30(6), R257-R258.
- Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A Practical Introduction*. Academic Press.
- Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial dependence across perception, attention, and memory. *Trends in Cognitive Sciences*, 21(7), 493-497.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, 36(14), 1.

- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12), 712-719.
- Koch, G., Oliveri, M., & Caltagirone, C. (2009). Neural networks engaged in milliseconds and seconds time processing: evidence from transcranial magnetic stimulation and patients with cortical or subcortical dysfunction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1525), 1907-1918.
- Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of neurophysiology*, 97(5), 3155-3164.
- Kotani, M., Kiyoshi, A., Murai, T., Nakako, T., Matsumoto, K., Matsumoto, A., ... & Ikeda, K. (2016). The dopamine D1 receptor agonist SKF-82958 effectively increases eye blinking count in common marmosets. *Behavioural brain research*, 300, 25-30.
- Kowalski, W. J. (1943). The effect of delay upon the duplication of short temporal intervals. *Journal of Experimental Psychology*, 33(3), 239.
- Körding, K. P., Ku, S. P., & Wolpert, D. M. (2004). Bayesian integration in force estimation. *Journal of neurophysiology*, 92(5), 3161-3165.
- Kraus, N., McGee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., & Koch, D. B. (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science*, 273(5277), 971-973.
- Lapid, E., Ulrich, R., & Rammsayer, T. (2008). On estimating the difference limen in duration discrimination tasks: A comparison of the 2AFC and the reminder task. *Perception & psychophysics*, 70(2), 291-305.
- Lacouture, Y., & Cousineau, D. (2008). How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. *Tutorials in quantitative methods for psychology*, 4(1), 35-45.
- Lamy, D., Alon, L., Carmel, T., & Shalev, N. (2015). The role of conscious perception in attentional capture and object-file updating. *Psychological science*, 26(1), 48-57.
- Lange, K. W., Tucha, O., Steup, A., Gsell, W., & Naumann, M. (1995). Subjective time estimation in Parkinson's disease. *Journal of neural transmission. Supplementum*, 46, 433-438.
- Lavoie, P., & Grondin, S. (2004). Information processing limitations as revealed by temporal discrimination. *Brain and cognition*, 54(3), 198-200.
- Lejeune, H. (1998). Switching or gating? The attentional challenge in cognitive models of psychological time. *Behavioural processes*, 44(2), 127-145.
- Lejeune, H. (2000). Prospective timing, attention and the switch: A response to 'Gating or switching? Gating is a better model of prospective timing' by Zakay. *Behavioural Processes*, 52(2-3), 71-76.
- Lejeune, H., & Wearden, J. H. (2006). Scalar properties in animal timing: Conformity and violations. *The Quarterly journal of experimental psychology*, 59(11), 1875-1908.
- Lejeune, H., & Wearden, J. H. (2009). Vierordt's The Experimental Study of the Time Sense (1868) and its legacy. *European Journal of Cognitive Psychology*, 21(6), 941-960.

- Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current biology*, 24(21), 2569-2574.
- Lieder, F., Daunizeau, J., Garrido, M. I., Friston, K. J., & Stephan, K. E. (2013). Modelling trial-by-trial changes in the mismatch negativity. *PLoS computational biology*, 9(2), e1002911.
- Linares, D., Cos, I., & Roseboom, W. (2016). Adaptation for multisensory relative timing. *Current Opinion in Behavioral Sciences*, 8, 35-41.
- Liu, W., Zhang, Z. J., Zhao, Y. J., Liu, Z. F., & Li, B. C. (2013). Effects of awareness on numerosity adaptation. *PloS One*, 8(10), e77556.
- Loewy, D. H., Campbell, K. B., & Bastien, C. (1996). The mismatch negativity to frequency deviant stimuli during natural sleep. *Electroencephalography and Clinical Neurophysiology*, 98(6), 493-501.
- Loveless, N. E., & Sanford, A. J. (1974). Slow potential correlates of preparatory set. *Biological psychology*, 1(4), 303-314.
- Luu, D. (2017). Keyboard latency. Dan Luu. Retrieved December 21, 2020, from <http://danluu.com/keyboard-latency/>
- Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual review of neuroscience*, 37, 205-220.
- Macar, F., Vidal, F., & Casini, L. (1999). The supplementary motor area in motor and sensory timing: evidence from slow brain potential changes. *Experimental brain research*, 125(3), 271-280.
- Macar, F., & Vidal, F. (2003). The CNV peak: an index of decision making and temporal memory. *Psychophysiology*, 40(6), 950-954.
- Macar, F., & Vidal, F. (2004). Event-related potentials as indices of time processing: a review. *Journal of Psychophysiology*, 18(2/3), 89-104.
- Macar, F., & Vitton, N. (1982). An early resolution of contingent negative variation (CNV) in time discrimination. *Electroencephalography and clinical neurophysiology*, 54(4), 426-435.
- MacKay, W. A., & Bonnet, M. (1990). CNV, stretch reflex and reaction time correlates of preparation for movement direction and force. *Electroencephalography and clinical Neurophysiology*, 76(1), 47-62.
- Malmi, R. A., & Samson, D. J. (1983). Intuitive averaging of categorized numerical stimuli. *Journal of Verbal Learning and Verbal Behavior*, 22(5), 547-559.
- Maricq, A. V., & Church, R. M. (1983). The differential effects of haloperidol and methamphetamine on time estimation in the rat. *Psychopharmacology*, 79(1), 10-15.
- Marinho, V., Oliveira, T., Rocha, K., Ribeiro, J., Magalhães, F., Bento, T., ... & Orsini, M. (2018). The dopaminergic system dynamic in the time perception: a review of the evidence. *International Journal of Neuroscience*, 128(3), 262-282.
- Mateeff, S., Hohnsbein, J., & Noack, T. (1985). Dynamic visual capture: apparent auditory motion induced by a moving visual target. *Perception*, 14(6), 721-727.

- Matell, M. S., King, G. R., & Meck, W. H. (2004). Differential modulation of clock speed by the administration of intermittent versus continuous cocaine. *Behavioral neuroscience*, 118(1), 150.
- Matell, M. S., & Meck, W. H. (2004). Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Cognitive brain research*, 21(2), 139-170.
- Maule, J., & Franklin, A. (2020). Adaptation to variance generalizes across visual domains. *Journal of Experimental Psychology: General*, 149(4), 662.
- May, P., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, 47(1), 66–122.
- May, P., Tiitinen, H., Ilmoniemi, R. J., Nyman, G., Taylor, J. G., & Näätänen, R. (1999). Frequency change detection in human auditory cortex. *Journal of computational neuroscience*, 6(2), 99-120.
- Meck, W. H. (1983). Selective adjustment of the speed of internal clock and memory processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9(2), 171.
- Meck, W. H. (1986). Affinity for the dopamine D2 receptor predicts neuroleptic potency in decreasing the speed of an internal clock. *Pharmacology Biochemistry and Behavior*, 25(6), 1185-1189.
- Meck, W. H. (1996). Neuropharmacology of timing and time perception. *Cognitive brain research*, 3(3-4), 227-242.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6), 1087-1092.
- Miall, C. (1989). The storage of time intervals using oscillating neurons. *Neural Computation*, 1(3), 359-371.
- Mikhael, J. G., & Gershman, S. J. (2019). Adapting the flow of time with dopamine. *Journal of neurophysiology*, 121(5), 1748-1760.
- Milne, E. (2011). Increased intra-participant variability in children with autistic spectrum disorders: Evidence from single-trial analysis of evoked EEG. *Frontiers in Psychology*, 2, 1–12.
- Mioni, G., Grassi, M., Tarantino, V., Stablum, F., Grondin, S., & Bisiacchi, P. S. (2016). The impact of a concurrent motor task on auditory and visual temporal discrimination tasks. *Attention, Perception, & Psychophysics*, 78(3), 742-748.
- Miyazaki, M., Nozaki, D., & Nakajima, Y. (2005). Testing Bayesian models of human coincidence timing. *Journal of neurophysiology*, 94(1), 395-399.
- Mohr, C., Sándor, P. S., Landis, T., Fathi, M., & Brugger, P. (2005). Blinking and schizotypal thinking. *Journal of Psychopharmacology*, 19(5), 513-520.
- Moradi, F., Koch, C., & Shimojo, S. (2005). Face adaptation depends on seeing the face. *Neuron*, 45(1), 169-175.

- Morgan, M. J. (2014). A bias-free measure of retinotopic tilt adaptation. *Journal of Vision*, 14(1), 7-7.
- Morgan, M. L., DeAngelis, G. C., & Angelaki, D. E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, 59(4), 662-673.
- Morgan, M., Chubb, C., & Solomon, J. A. (2008). A 'dipper' function for texture discrimination based on orientation variance. *Journal of vision*, 8(11), 9-9.
- Morgan, M. J., & Glennerster, A. (1991). Efficiency of locating centres of dot-clusters by human observers. *Vision research*, 31(12), 2075-2083.
- Morris, B. J., & Masnick, A. M. (2015). Comparing data sets: Implicit summaries of the statistical properties of number sets. *Cognitive science*, 39(1), 156-170.
- Morris, J. P., Pelphrey, K. A., & McCarthy, G. (2007). Controlled scanpath variation alters fusiform face activation. *Social Cognitive and Affective Neuroscience*, 2(1), 31-38.
- Moutoussis, K., & Zeki, S. (2002). The relationship between cortical activation and perception investigated with invisible stimuli. *Proceedings of the National Academy of Sciences*, 99(14), 9527-9532.
- Muller-Gass, A., Duncan, M., Tavakoli, P., & Campbell, K. (2019). Individual differences in feedback processing affect perceptual learning. *Personality and Individual Differences*, 143, 145-154.
- Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal*, 24(3), 69-71.
- Murai, Y., & Yotsumoto, Y. (2016). Timescale-and sensory modality-dependency of the central tendency of time perception. *PloS one*, 11(7).
- Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & psychophysics*, 70(5), 772-788.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology*, 118(12), 2544-2590.
- Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clinical neurophysiology*, 115(1), 140-144.
- Ng, K. K., Tobin, S., & Penney, T. B. (2011). Temporal accumulation and decision processes in the duration bisection task revealed by contingent negative variation. *Frontiers in integrative neuroscience*, 5, 77.
- Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2015). Direct encoding of orientation variance in the visual system. *Journal of vision*, 15(4), 3-3.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Contestabile, M. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425.

- O'Shea, R. P., & Crassini, B. (1981). Interocular transfer of the motion after-effect is not reduced by binocular rivalry. *Vision research*, 21(6), 801-804.
- Oliveri, M., Vicario, C. M., Salerno, S., Koch, G., Turriziani, P., Mangano, R., Chillemi, G. & Caltagirone, C. 2008 Perceiving numbers alters time perception. *Neurosci. Lett.* 438, 308–311.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251).
- Ornstein, D. S. (1969). Random walks. I. *Transactions of the American Mathematical Society*, 138, 1-43.
- Ostwald, D., Spitzer, B., Guggenmos, M., Schmidt, T. T., Kiebel, S. J., & Blankenburg, F. (2012). Evidence for neural encoding of Bayesian surprise in human somatosensation. *NeuroImage*, 62(1), 177-188.
- Overholser, B. R., & Sowinski, K. M. (2008). Biostatistics primer: part 2. *Nutrition in clinical practice*, 23(1), 76-84.
- Papadimitriou, C., Ferdoash, A., & Snyder, L. H. (2015). Ghosts in the machine: memory interference from the previous trial. *Journal of neurophysiology*, 113(2), 567-577.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?. *Perspectives on Psychological Science*, 7(6), 528-530.
- Pastor, M. A., Artieda, J., Jahanshahi, M., & Obeso, J. A. (1992). Time estimation and reproduction is abnormal in Parkinson's disease. *Brain*, 115(1), 211-225.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature neuroscience*, 4(7), 739-744
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Richard H"ochenberger, Sogo, H., ... Jonas Kristoffer Lindelov. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203.
- Petzschnner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: a study on human path integration. *Journal of Neuroscience*, 31(47), 17220-17229.
- Petzschnner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in cognitive sciences*, 19(5), 285-293.
- Penton-Voak, I. S., Edwards, H., Percival, A., & Wearden, J. H. (1996). Speeding up an internal clock in humans? Effects of click trains on subjective duration. *Journal of Experimental Psychology: Animal Behavior Processes*, 22(3), 307.
- Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological science*, 24(8), 1389-1397.
- Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *arXiv preprint arXiv:1901.07915*.

- Plant, R. R. (2016). A reminder on millisecond timing accuracy and potential replication failure in computer-based psychology experiments: An open letter. *Behavior Research Methods*, 48(1), 408-411.
- Plant, R. R., & Quinlan, P. T. (2013). Could millisecond timing errors in commonly used equipment be a cause of replication failure in some neuroscience studies?. *Cognitive, Affective, & Behavioral Neuroscience*, 13(3), 598-614.
- Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature neuroscience*, 16(9), 1170-1178.
- Pouthas, V., George, N., Poline, J. B., Pfeuty, M., VandeMoortele, P. F., Hugueville, L., ... & Renault, B. (2005). Neural network involved in time perception: an fMRI study comparing long and short interval estimation. *Human brain mapping*, 25(4), 433-441.
- Penton-Voak, I. S., Edwards, H., Percival, A., & Wearden, J. H. (1996). Speeding up an internal clock in humans? Effects of click trains on subjective duration. *Journal of Experimental Psychology: Animal Behavior Processes*, 22(3), 307.
- Power, R. P., & Graham, A. (1976). Dominance of touch by vision: generalization of the hypothesis to a tactually experienced population. *Perception*, 5(2), 161-166.
- Prins, N., & Kingdon, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. *Palamedes: Matlab routines for analyzing psychophysical data*.
- Rammsayer, T. (1989a). Dopaminergic and serotonergic influence on duration discrimination and vigilance. *Pharmacopsychiatry*, 22(S 1), 39-43.
- Rammsayer, T. (1989b). Is there a common dopaminergic basis of time perception and reaction time?. *Neuropsychobiology*, 21(1), 37-42.
- Rammsayer, T. (1990). Temporal discrimination in schizophrenic and affective disorders: evidence for a dopamine-dependent internal clock. *International Journal of Neuroscience*, 53(2-4), 111-120.
- Rammsayer, T. H. (1999). Neuropharmacological evidence for different timing mechanisms in humans. *The Quarterly Journal of Experimental Psychology: Section B*, 52(3), 273-286.
- Rammsayer, T. H. (2014). The effects of type of interval, sensory modality, base duration, and psychophysical task on the discrimination of brief time intervals. *Attention, Perception, & Psychophysics*, 76(4), 1185-1196.
- Rammsayer, T., & Pichelmann, S. (2018). Visual-auditory differences in duration discrimination depend on modality-specific, sensory-automatic temporal processing: Converging evidence for the validity of the Sensory-Automatic Timing Hypothesis. *Quarterly Journal of Experimental Psychology*.
- Randeniya, R., Oestreich, L., & Garrido, M. (2019). Auditory prediction errors and auditory white matter microstructure as predictors of psychotic experiences in healthy individuals. *bioRxiv*.
- Regan, D., & Tansley, B. W. (1979). Selective adaptation to frequency-modulated tones: Evidence for an information-processing channel selectively sensitive to frequency changes. *The Journal of the Acoustical Society of America*, 65(5), 1249-1257.

- Ren, Y., Allenmark, F., Müller, H. J., & Shi, Z. (2020). Logarithmic encoding of ensemble time intervals. *Scientific Reports*, 10.
- Reuschel, J., Drewing, K., Henriques, D. Y., Rösler, F., & Fiehler, K. (2010). Optimal integration of visual and proprioceptive movement information for the perception of trajectory geometry. *Experimental brain research*, 201(4), 853-862.
- Rhodes, D., Seth, A. K., & Roseboom, W. (2018). Multiple Duration Priors Within and Across the Senses. *bioRxiv*, 467027.
- Riesen, J. M., & Schnider, A. (2001). Time estimation in Parkinson's disease: normal long duration estimation despite impaired short duration discrimination. *Journal of neurology*, 248(1), 27-35.
- Roach, N. W., McGraw, P. V., Whitaker, D. J., & Heron, J. (2017). Generalization of prior information for rapid Bayesian time estimation. *Proceedings of the National Academy of Sciences*, 114(2), 412-417.
- Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, 143(3606), 594-596
- Roger, C., Hasbroucq, T., Rabat, A., Vidal, F., & Burle, B. (2009). Neurophysics of temporal discrimination in the rat: a mismatch negativity study. *Psychophysiology*, 46(5), 1028-1032.
- Roseboom, W. (2019). Serial dependence in timing perception. *Journal of experimental psychology: human perception and performance*, 45(1), 100.
- Roseboom, W., Fountas, Z., Nikiforou, K., Bhowmik, D., Shanahan, M., & Seth, A. K. (2019). Activity in perceptual classification networks as a basis for human subjective time perception. *Nature communications*, 10(1), 267.
- Roseboom, W., Linares, D., & Nishida, S. Y. (2015). Sensory adaptation for timing perception. *Proceedings of the Royal Society B: Biological Sciences*, 282(1805), 20142833.
- Rousseau, R., & Kristofferson, A. B. (1973). The discrimination of bimodal temporal gaps. *Bulletin of the Psychonomic Society*, 1(2), 115-116.
- Roumani, D., & Moutoussis, K. (2020). Invisible Adaptation: The Effect of Awareness on the Strength of the Motion Aftereffect. *Perception*, 49(8), 835-857.
- Rousseu, R., Poirier, J., & Lemyre, L. (1983). Duration discrimination of empty time intervals marked by intermodal pulses. *Perception & Psychophysics*, 34(6), 541-548.
- Ruchkin, D. S., Sutton, S., Mahaffey, D., & Glaser, J. (1986). Terminal CNV in the absence of motor response. *Electroencephalography and clinical neurophysiology*, 63(5), 445-463.
- Ryan, L. J. (2011). Temporal context affects duration reproduction. *Journal of Cognitive Psychology*, 23(1), 157-170
- Saarinén, J., Paavilainen, P., Schöger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *NeuroReport*, 3(12), 1149-1151.

- Sandberg, K., & Overgaard, M. (2015). Using the perceptual awareness scale (PAS). *Behavioural methods in consciousness research*, 181-195.
- Santesso, D. L., Dillon, D. G., Birk, J. L., Holmes, A. J., Goetz, E., Bogdan, R., & Pizzagalli, D. A. (2008). Individual differences in reinforcement learning: behavioral, electrophysiological, and neuroimaging correlates. *Neuroimage*, 42(2), 807-816.
- Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, 56(6), e13335.
- Schweickert, R., Han, H. J., Yamaguchi, M., & Fortin, C. (2014). Estimating averages from distributions of tone durations. *Attention, Perception, & Psychophysics*, 76(2), 605-620.
- Sherman, M. T., Fountas, Z., Seth, A. K., & Roseboom, W. (2020). Accumulation of Salient Perceptual Events Predicts Subjective Time. *bioRxiv*
- Shi, Z., & Burr, D. (2016). Predictive coding of multisensory timing. *Current Opinion in Behavioral Sciences*, 8, 200-206.
- Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends in cognitive sciences*, 17(11), 556-564.
- Skagerlund, K., Karlsson, T., & Träff, U. (2016). Magnitude processing in the brain: an fMRI study of time, space, and numerosity as a shared cortical system. *Frontiers in Human Neuroscience*, 10, 500.
- Smith, J. G., Harper, D. N., Gittings, D., & Abernethy, D. (2007). The effect of Parkinson's disease on time estimation as a function of stimulus duration range and modality. *Brain and cognition*, 64(2), 130-143.
- Soares, S., Atallah, B. V., & Paton, J. J. (2016). Midbrain dopamine neurons control judgment of time. *Science*, 354(6317), 1273-1277.
- Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of vision*, 11(12), 13-13.
- Sokolowski, H. M., Fias, W., Ononye, C. B., & Ansari, D. (2017). Are numbers grounded in a general magnitude processing system? A functional neuroimaging meta-analysis. *Neuropsychologia*, 105, 50-69.
- Spellman, B. A., Gilbert, E. A., & Corker, K. S. (2018). Open science. *Stevens' handbook of experimental psychology and cognitive neuroscience*, 5, 1-47.
- Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Frontiers in human neuroscience*, 8, 666.
- Stevens, S. S., & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Perception & Psychophysics*, 1(5), 439-446.
- Stocker, A. A., & Simoncelli, E. P. (2006a). Sensory adaptation within a Bayesian framework for perception. In *Advances in neural information processing systems* (pp. 1289-1296).
- Stocker, A. A., & Simoncelli, E. P. (2006b). Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience*, 9(4), 578-585.

- Strömmer, J. M., Tarkka, I. M., & Astikainen, P. (2014). Somatosensory mismatch response in young and elderly adults. *Frontiers in aging neuroscience*, 6, 293.
- Suárez-Pinilla, M. (2018). *Contextual modulation of visual variability: perceptual biases over time and across the visual field* (Doctoral dissertation, University of Sussex).
- Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018). Serial dependence in the perception of visual variance. *Journal of vision*, 18(7), 4-4.
- Suárez-Pinilla, M., Nikiforou, K., Fountas, Z., Seth, A. K., Roseboom, W., Zwaan, R., & Zacks, J. (2019). Perceptual content, not physiological signals, determines perceived duration when viewing dynamic, natural scenes. *Collabra: Psychology*, 5(1).
- Symonds, R. M., Lee, W. W., Kohn, A., Schwartz, O., Witkowski, S., & Sussman, E. S. (2017). Distinguishing neural adaptation and predictive coding hypotheses in auditory change detection. *Brain topography*, 30(1), 136-148.
- Teghtsoonian, R., & Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics*, 24(4), 305-314.
- Terhune, D. B., Sullivan, J. G., & Simola, J. M. (2016). Time dilates after spontaneous blinking. *Current Biology*, 26(11), R459-R460.
- Tokita, M., Ueda, S., & Ishiguchi, A. (2016). Evidence for a global sampling process in extraction of summary statistics of item sizes in a set. *Frontiers in psychology*, 7, 711.
- Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychological Monographs: General and Applied*, 77(13), 1.
- Tse, C. Y., & Penney, T. B. (2006). Preattentive timing of empty intervals is from marker offset to onset. *Psychophysiology*, 43(2), 172-179.
- Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature neuroscience*, 8(8), 1096-1101.
- Van Beers, R. J., Sittig, A. C., & Gon, J. J. D. V. D. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of neurophysiology*, 81(3), 1355-1364.
- Van der Post, J., De Waal, P. P., De Kam, M. L., Cohen, A. F., & van Gerven, J. M. A. (2004). No evidence of the usefulness of eye blinking as a marker for central dopaminergic activity. *Journal of Psychopharmacology*, 18(1), 109-114.
- VanderWerf, F., Brassinga, P., Reits, D., Aramideh, M., & Ongerboer de Visser, B. (2003). Eyelid movements: behavioral studies of blinking in humans under different stimulus conditions. *Journal of neurophysiology*, 89(5), 2784-2796.
- Vandenbroucke, A. R., Fahrenfort, J. J., Meuwese, J. D. I., Scholte, H. S., & Lamme, V. A. F. (2014). Prior knowledge about objects determines neural color representation in human visual cortex. *Cerebral cortex*, 26(4), 1401-1408.
- Vidal, F., Bonnet, M., & Macar, F. (1995). Programming the duration of a motor sequence: role of the primary and supplementary motor areas in man. *Experimental Brain Research*, 106(2), 339-350.

- Vilidaite, G., Yu, M., & Baker, D. H. (2017). Internal noise estimates correlate with autistic traits. *Autism Research*, 10(8), 1384-1391.
- Volkman, F. C., Riggs, L. A., & Moore, R. K. (1980). Eyeblinks and visual suppression. *Science*, 207(4433), 900-902.
- Vroomen, J., & de Gelder, B. (2004). Temporal ventriloquism: sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance*, 30(3), 513
- Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, 32(11), 3665-3678.
- Wade, N. J., & Wenderoth, P. (1978). The influence of colour and contour rivalry on the magnitude of the tilt after-effect. *Vision research*, 18(7), 827-835.
- Wahl, O. F., & Sieg, D. (1980). Time estimation among schizophrenics. *Perceptual and Motor Skills*.
- Walker, E. Y., Cotton, R. J., Ma, W. J., & Tolia, A. S. (2018). A neural basis of probabilistic computation in visual cortex. *bioRxiv*, 365973.
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in cognitive sciences*, 7(11), 483-488.
- Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: an electric sign of sensori-motor association and expectancy in the human brain. *Nature*, 203(4943), 380-384.
- Wang, J., Narain, D., Hosseini, E. A., & Jazayeri, M. (2018). Flexible timing by temporal scaling of cortical responses. *Nature neuroscience*, 21(1), 102-110.
- Wang, L., Rangarajan, K. V., Gerfen, C. R., & Krauzlis, R. J. (2018). Activation of striatal neurons causes a perceptual decision bias during visual change detection in mice. *Neuron*, 97(6), 1369-1381.
- Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision research*, 32(5), 931-941.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics*, 33(2), 113-120.
- Wearden, J. H. (1991). Do humans possess an internal clock with scalar timing properties?. *Learning and Motivation*, 22(1-2), 59-83.
- Wearden, J. H. (2003). Applying the scalar timing model to human time psychology: Progress and challenges. *Time and mind II: Information processing perspectives*, 21-39.
- Wearden, J. H., & McShane, B. (1988). Interval production as an analogue of the peak procedure: Evidence for similarity of human and animal timing processes. *The Quarterly Journal of Experimental Psychology*, 40(4), 363-375.
- Webster, M. A. (2011). Adaptation and visual coding. *Journal of vision*, 11(5), 3-3.
- Webster, M. A. (2015). Visual adaptation. *Annual review of vision science*, 1, 547-567.

- Webster, M. A., & MacLeod, D. I. (2011). Visual adaptation and face perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1571), 1702-1725.
- Wei, X. X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nature neuroscience*, 18(10), 1509.
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological bulletin*, 88(3), 638.
- Wenderoth, P., & Wiese, M. (2008). Retinotopic encoding of the direction aftereffect. *Vision research*, 48(19), 1949-1954.
- Wiener, M., & Thompson, J. C. (2015). Repetition enhancement and memory effects for duration. *Neuroimage*, 113, 268-278.
- Whitaker, D., McGraw, P. V., Pacey, I., & Barrett, B. T. (1996). Centroid analysis predicts visual localization of first-and second-order stimuli. *Vision Research*, 36(18), 2957-2970.
- Whiting, B. F., & Oriet, C. (2011). Rapid averaging? Not so fast!. *Psychonomic Bulletin & Review*, 18(3), 484-489.
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual review of psychology*, 69, 105-129.
- Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *ACM SIGGRAPH Computer Graphics*, 18(1), 24-24.
- Wilkinson, R. T., & Allison, S. (1989). Age and simple reaction time: decade differences for 5,325 subjects. *Journal of gerontology*, 44(2), 29-35.
- Wimmer, R., Schmid, A., & Bockes, F. (2019, May). On the Latency of USB-Connected Input Devices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).
- Winkler, I. (2007). Interpreting the mismatch negativity. *Journal of Psychophysiology*, 21(3-4), 147-163.
- Winkler, I., & Czigler, I. (1998). Mismatch negativity: deviance detection or the maintenance of the 'standard'. *NeuroReport*, 9(17), 3809-3813.
- Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain research*, 742(1-2), 239-252.
- Winter, O., Kok, A., Kennermans, J. L., & Elton, M. (1995). Auditory event-related potentials to deviant stimuli during drowsiness and stage 2 sleep. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 96(5), 398-412.
- Wise, R. A. (2004). Dopamine, learning and motivation. *Nature reviews neuroscience*, 5(6), 483-494.
- Withagen, R., & Van Wermeskerken, M. (2009). Individual differences in learning to perceive length by dynamic touch: Evidence for variation in perceptual learning capacities. *Perception & Psychophysics*, 71(1), 64-75.

- Woodrow, H. (1930). The reproduction of temporal intervals. *Journal of Experimental Psychology*, 13(6), 473.
- Woodrow, H. (1934). The temporal indifference interval determined by the method of mean error. *Journal of Experimental Psychology*, 17(2), 167.
- Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of vision*, 7(10), 2-2.
- Yang, E., Hong, S. W., & Blake, R. (2010). Adaptation aftereffects to facial expressions suppressed from visual awareness. *Journal of Vision*, 10(12), 24-24.
- Yates, M. J., Loetscher, T., & Nicholls, M. E. (2012). A generalized magnitude system for space, time, and quantity? A cautionary note. *Journal of Vision*, 12(7), 9-9.
- Zakay, D. (2000). Gating or switching? Gating is a better model of prospective timing (a response to 'switching or gating?' by Lejeune). *Behavioural processes*, 52(2-3), 63-69.
- Zakay, D., & Block, R. A. (1996). The role of attention in time estimation processes. *In Advances in psychology* (Vol 115, pp. 143-164). North-Holland.
- Zeiler, M. D., Scott, G. K., & Hoyert, M. S. (1987). Optimal temporal differentiation. *Journal of the experimental analysis of behavior*, 47(2), 191-200.

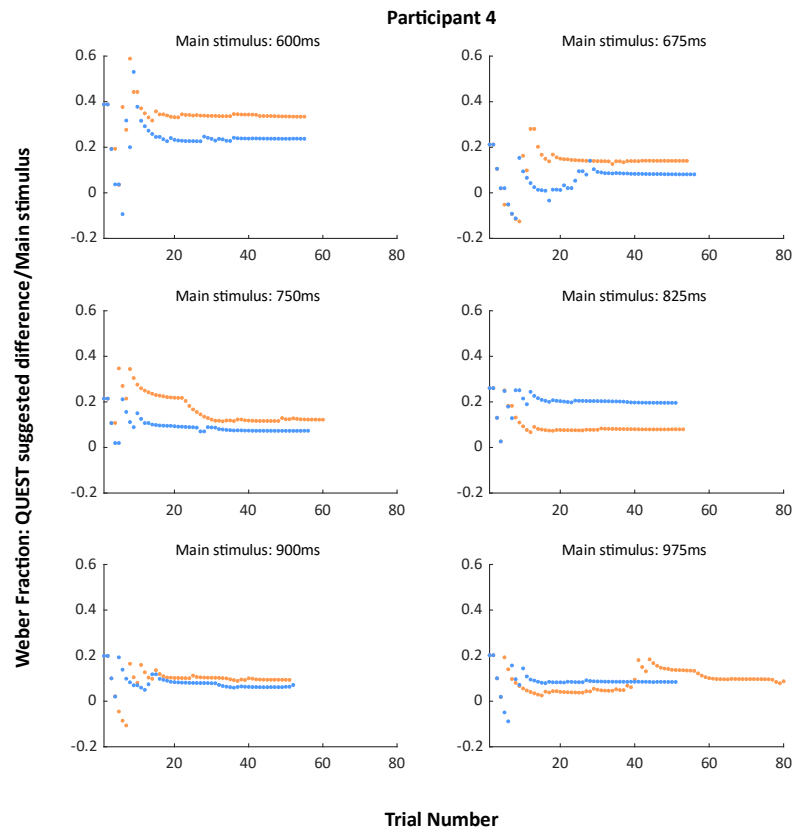
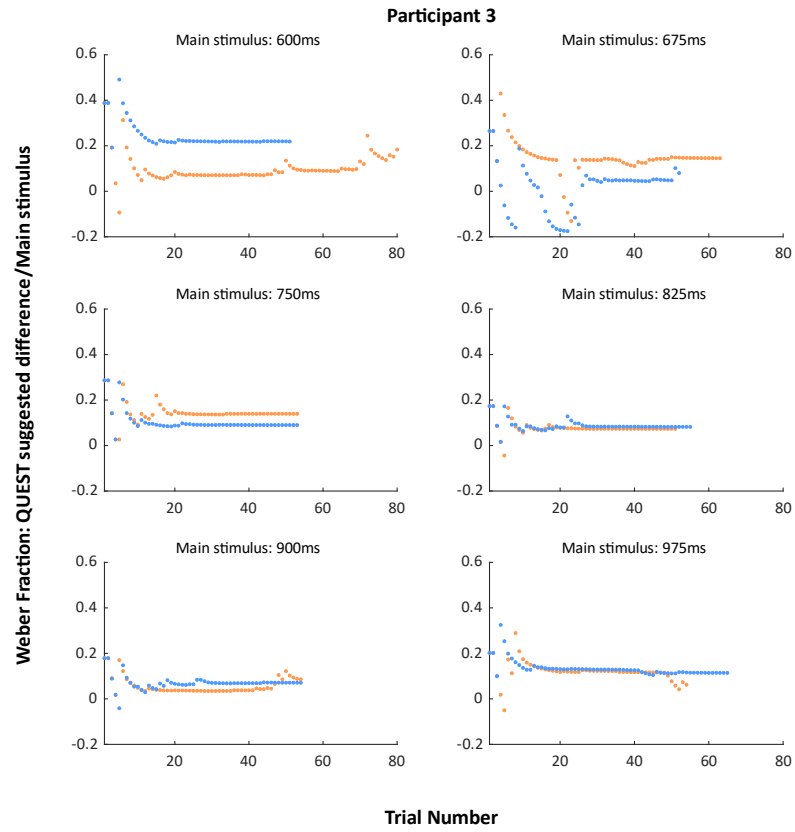
Appendix

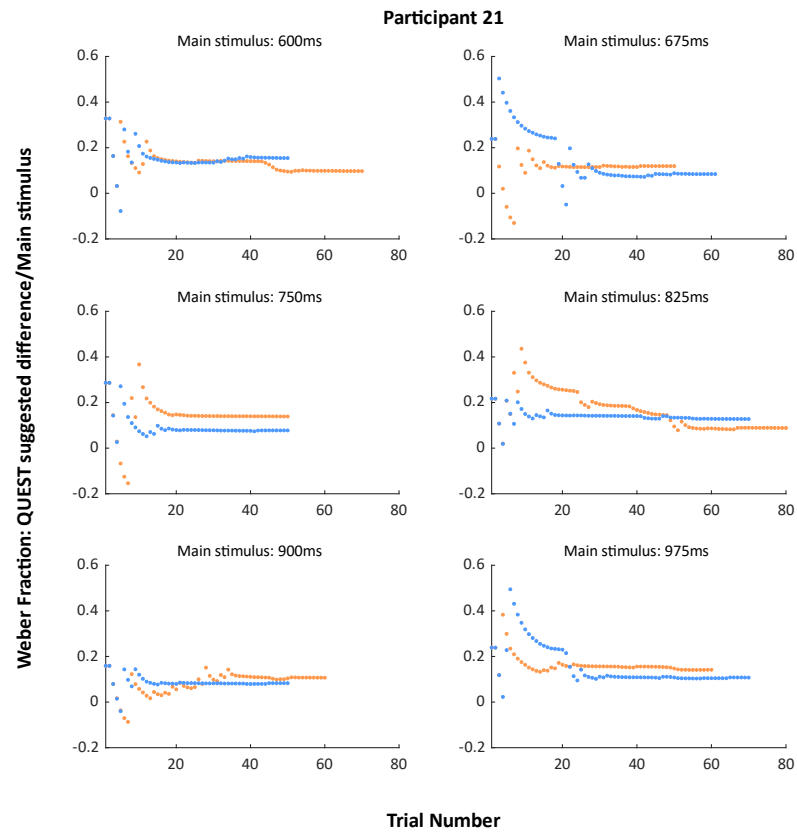
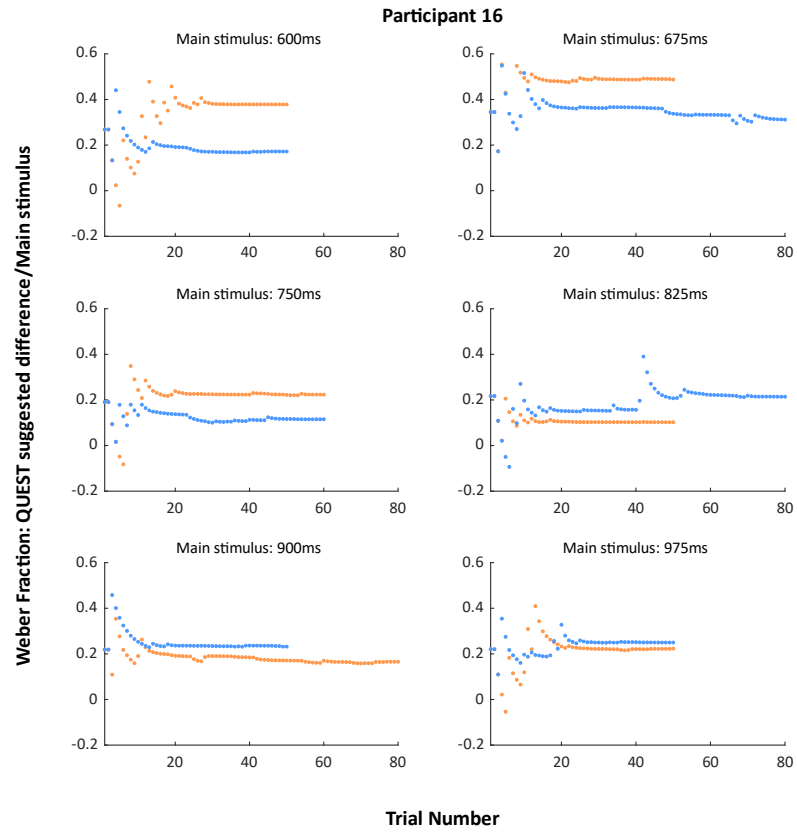
Appendix A. 2IFC threshold convergence and distribution, Chapter 4

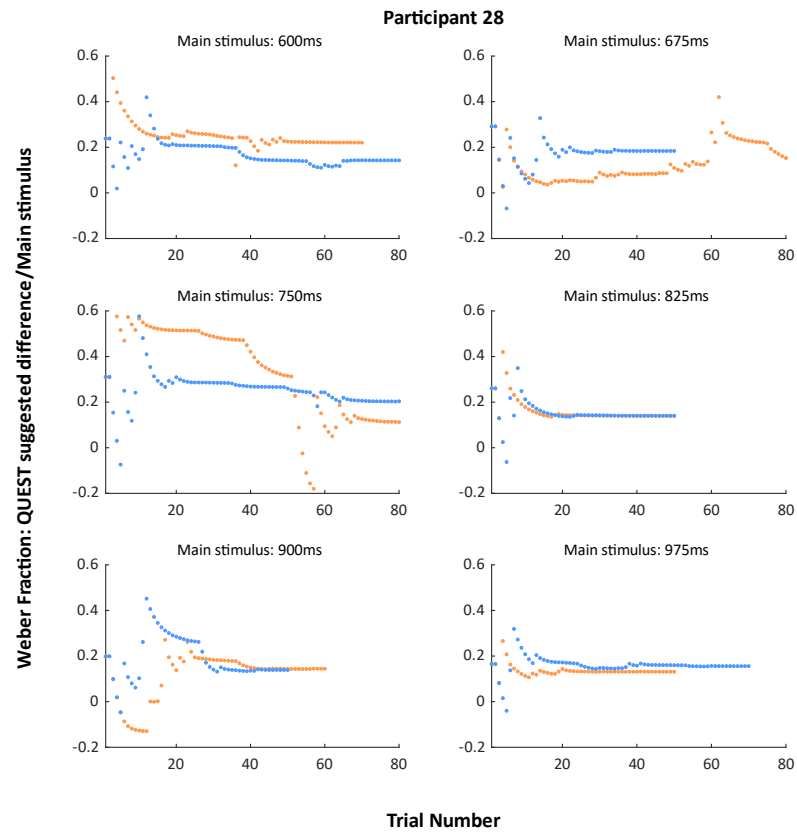
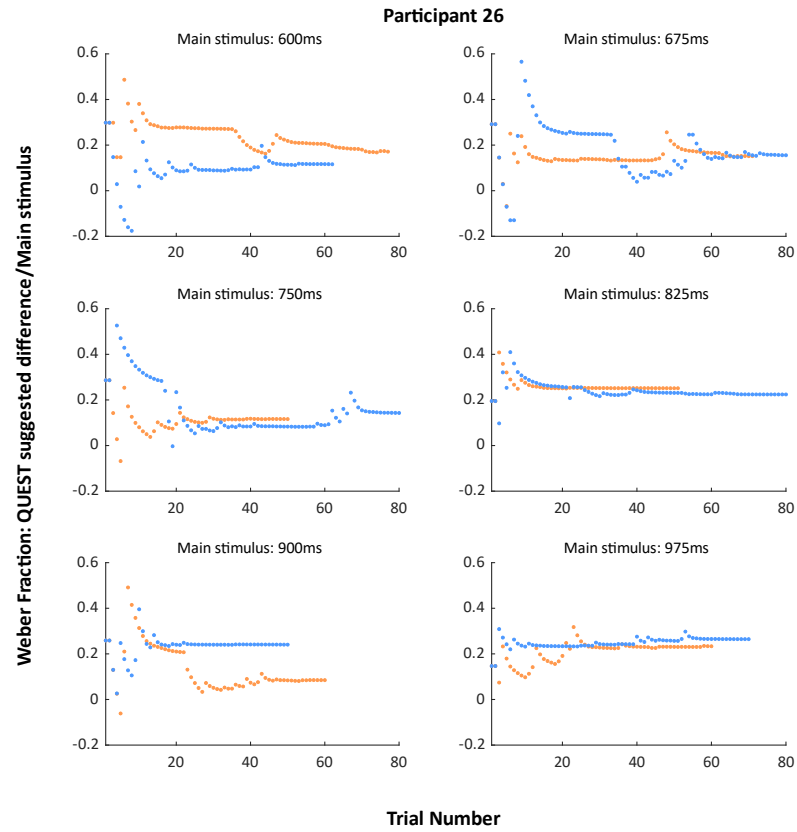
Table 1A. Convergence of the 75% threshold (in ms) for each participant, stimulus, and block. Cells highlighted in green signify that the criterion for convergence was met. Cells highlighted in orange signify that the criterion for convergence was not met.

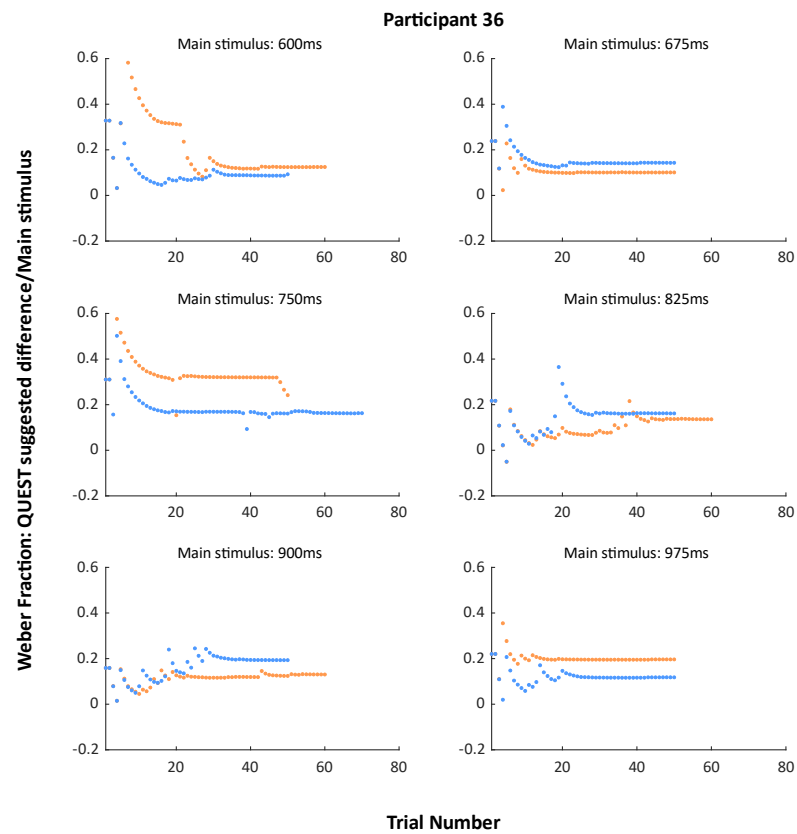
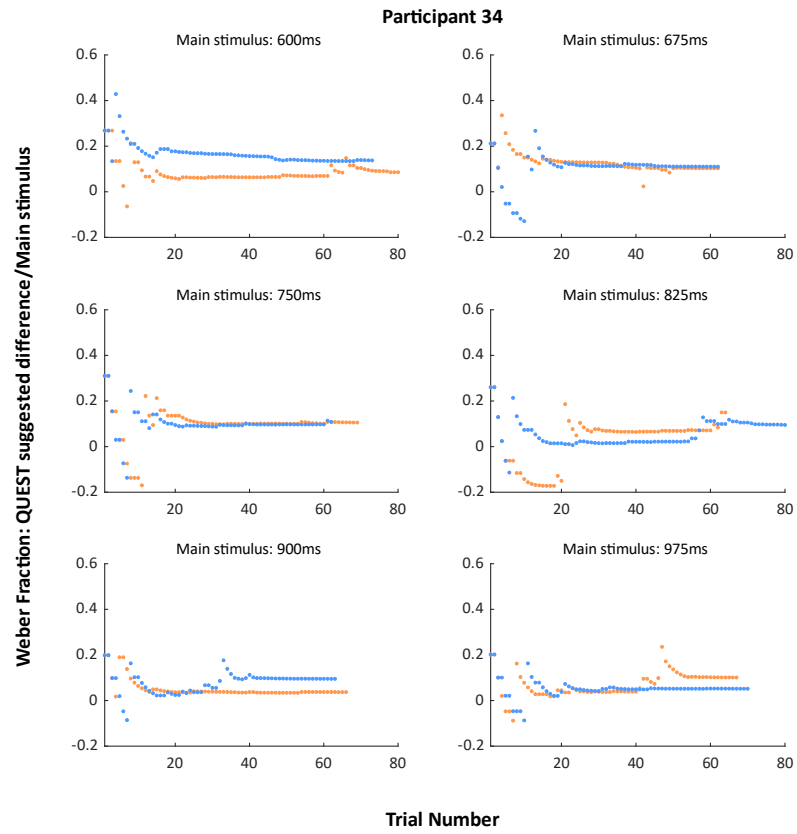
Participant ID	Main stimulus (ms)											
	600		675		750		825		900		975	
	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2
3	60.67	67.76	77.31	63.86	97.62	46.15	53.47	111.56	104.61	67.60	110.91	127.08
4	94.79	54.76	91.62	55.06	81.85	61.88	93.90	93.90	65.69	161.47	200.52	142.37
16	167.61	86.08	226.93	102.96	84.31	176.84	149.07	208.74	217.54	243.61	328.86	211.57
21	95.24	74.59	80.50	57.01	73.16	105.33	58.42	92.73	102.99	60.62	138.11	104.64
26	227.08	257.99	206.24	184.91	102.82	69.54	102.29	102.30	87.01	107.60	76.06	216.54
28	132.49	85.57	128.04	151.92	130.02	124.76	100.88	124.47	115.77	115.59	83.34	152.75
34	69.11	74.39	52.22	88.03	33.37	85.50	98.44	50.57	92.87	79.12	78.88	95.52
36	117.96	173.96	68.15	96.74	112.20	133.21	75.19	55.04	192.63	114.88	166.11	122.09
44	64.79	86.92	125.95	63.84	66.66	50.40	73.31	71.01	129.47	81.96	69.70	94.95

Figure A1: Trial-by-trial Weber fractions suggested by QUEST individually for each participant and stimulus across the two experiments in Chapter 4. On each trial of the 2IFC task, QUEST provided a suggestion for the difference between the main and comparison stimulus that should be tested next based on the participant's performance on the previous trials. Sometimes QUEST would suggest a difference smaller than 0 in which case the difference suggested by QUEST did not correspond to the actual difference in that trial (which was drawn from a uniform distribution between 0 and 10ms). In each block, QUEST convergence was assessed when participants completed 50 trials. If the last 10 trials in the block contained 5 consecutive trials between which the difference in the values suggested by QUEST was less than 0.5ms, we concluded that the estimate of the threshold has converged, and the block ended. If the difference between the suggested values was larger, participants completed additional chunks of 10 more trials each at the end of which the convergence was evaluated again. The maximum length of each block was 80 trials irrespective of whether the QUEST converged. Participants completed 2 blocks per stimulus. Trials from the first block are depicted in orange, trials from the second block are depicted in blue.









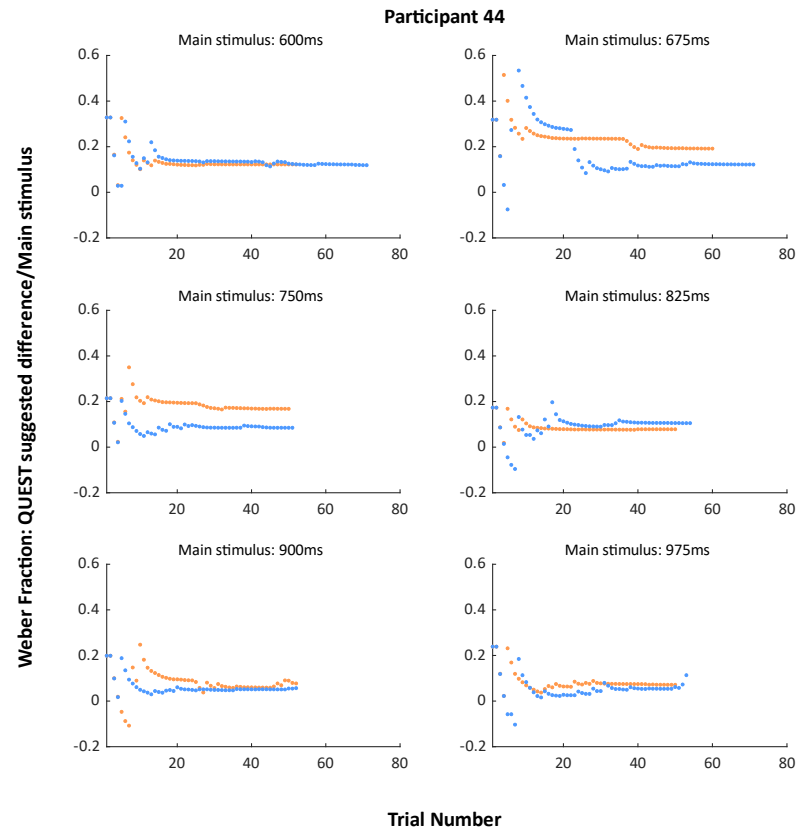
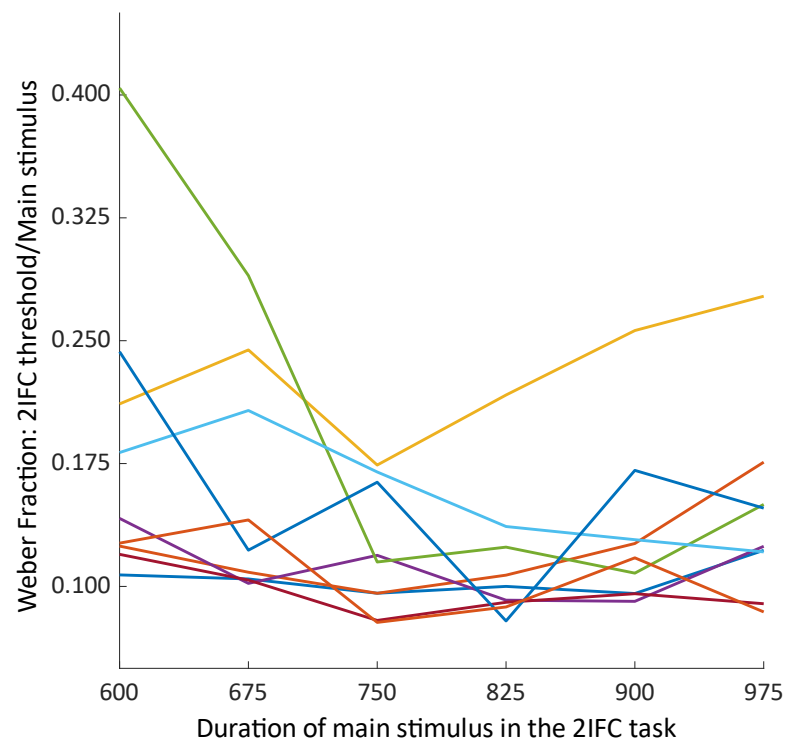


Figure A2: Average Weber fractions for each participant and each stimulus across the two experiments in Chapter 4. Each coloured line represented a different participant (N=9).



Appendix B. Instructions for tasks in Chapter 4

1. Duration reproduction

In this task, you will be presented with temporal intervals, and you will be asked to reproduce them. When a white fixation cross will appear on the screen, you can start the next trial. Keep looking at the fixation cross throughout the entire trial. Each trial will start with the presentation of a temporal interval. The start and end of the interval will be designated by two short events. The start of the interval will be a button click that you generate by briefly clicking the button on the button box. After some time, a black-and-white grating will appear on the screen for a very short time. The end of the interval will be signalled by the disappearance of the grating. It is important that you do not use any techniques to time or count how long the intervals are. All durations in the experiment are very short and difficult to count. Just pay attention to how much time passes between your button click and the disappearance of the grating. Shortly after the grating disappears, the fixation cross will turn green. This is your cue that you can provide a response. You will respond by reproducing the interval presented on the trial - keep the button on the button box pressed for as long as you think the interval between the click you generated, and the disappearance of the grating was. After providing your response, you will receive feedback [here participants were shown the feedback display]. The coloured line represents how your response was from the true duration of the interval. If the coloured line is exactly over the white line in the middle of the box, this means there was no error between your response and the presented interval. If the coloured line is to the left of the white line, it means your response was shorter than the presented interval. If the colour line is to the right of the white line, it means your response was longer than the presented interval. The further the coloured line is from the middle, the bigger the difference between your response and the stimulus. The colour of the coloured line also provides you with some information – if the line is green, it means your response was within 10% of the presented duration. If the line is brown, it means your response was further away. Your goal is to try to be as accurate as possible.

There are a few instances in which trials will be excluded and repeated later on in the block. If you start reproducing the interval before the response cue appears, the trial will be excluded, and a message saying “Too soon” will appear on the screen. If you start reproducing the interval longer than 1.5s after the response cue appears, the trial will be excluded, and a message saying “Too late” will appear on the screen. Trials will also be excluded and repeated if an error occurs during the response, in which case a message saying “Invalid trial” will appear on the screen.

The task will start with a training block containing 60 trials. In the training block, you will be presented with a sequence consisting of 10 repetitions of a short stimulus, a medium stimulus, and a long stimulus. They will be presented in ascending order – first all short stimuli, then all medium stimuli, and finally all long stimuli. The sequence will be presented twice. In the training block, you are not allowed to skip trials, and the feedback display is going to stay on the screen until you press the spacebar. These durations do not appear in the main testing blocks.

After the training block is over, you will complete a variable number of test blocks depending on your performance. Each test block takes around 40 minutes to complete, and it is divided into six shorter runs of ~100 trials each. The stimuli in the

test blocks will be between 600 and 975ms long and they will be shuffled. In the test blocks, when it is time to provide a response, you are allowed to skip the trial by pressing “Z”. Only skip trials if something happened that prevents you from estimating the duration of the interval was – for example, if you blinked when the grating was presented and you didn’t see the grating, or if you felt that you were not focused during the trial. Skipped trials will be presented later on in the block, so do not skip trials if they were difficult because they will be repeated. If you decide to skip a trial, a message saying “Trial skipped” will be displayed on the screen. Finally, in the test blocks, the feedback display will be presented for a short amount of time, and it will disappear on its own. The task is self-paced so you can take breaks whenever you need to. The task is difficult and requires concentration, so if you are tired it is best to take a break.

2. Duration discrimination

In this task, you will be presented with two temporal intervals one after the other, and you will be asked to report which interval was shorter. The start and end of each of the two intervals will be defined by a short event. The start of each interval will be a button click that you generate by briefly clicking the button on the button box. After some time, a black-and-white grating will appear on the screen for a very short time. The end of the interval will be signalled by the disappearance of the grating. Each trial will begin with the presentation of a white fixation cross – keep looking at the cross during the entire trial. The presentation of the cross is your cue that you can click the button to generate the first interval. After some time, a grating is going to appear on the screen – the disappearance of the grating is the end of the first interval. Shortly after that, the fixation cross will disappear. When the fixation cross appears again, this is your cue that you can generate the second interval. Click on the button box to start the second interval, after some time a grating will appear on the screen for a very short time. The disappearance of the grating signals the end of the second interval. Do not use any techniques to time the duration of either of the intervals. After some time, you will be asked to report which interval was shorter. Press “1” if think that the first interval was shorter, or “2” if you think the second interval was shorter. At this point, you can change your response. If at any point in the trial something happened that prevented you from paying attention to the intervals, you can skip the trial by pressing “Z”. For example, skip the trial if you blinked and didn’t see either of the gratings or if you felt that you were not focused on the task. Skipped trials will be repeated later in the block. When you are happy with your choice, press the spacebar to save your response. If you decided to skip the trial, a message saying “Trial skipped” will be presented on the screen. If you provided a response, a message will be presented stating whether your response was correct or incorrect.

You will complete 12 blocks, each block will be between 50-80 trials long, or ~5-10 minutes. Take a break at the end of each block, and in addition, you can take a break whenever you are tired within a block.

3. Reaction time task

In this task, you will be asked to keep the button on the button box pressed for some time and release it as fast as possible when the grating on the screen disappears.

You can start a trial when the white fixation cross appears on the screen. In this task, you have to press the button box and keep it pressed. When you press the button, a grating will appear on the screen and it will remain there for some time. Keep the button pressed. When the grating disappears, release the button as fast as you can. It is important to wait until the grating disappears before you initiate the release of the button. If you release the button before the grating disappears or too soon after the grating disappears, this will be considered a false-start, and the trial will be excluded and repeated. In this case, a message saying “Too soon” will appear on the screen.

You will complete 6 blocks, each block is 100 trials long, and should take ~5 minutes. Take a break at the end of each block, and if you are tired at any point within each block as well.

Appendix C. Testing robustness to parameter misspecification, Chapter 4

This section is included as Supplementary information in:

Baykova, R., Buckley, C., Seth, A., & Roseboom, W. (in-principle acceptance). Influence of prior distribution variance on time perception (Registered Report Stage 1). *Royal Society Open Science*. Available at: <https://osf.io/75p42/>

The Bayesian ideal observer model used in Chapter 4 (Model 3) may fail to generate meaningful predictions if the tasks we have designed to measure perceptual and motor noise do not provide truthful representations of the parameters that influence participants' performance in the duration reproduction task. Therefore, we conducted additional simulations to ensure that our proposed analysis is robust to parameter misspecification. Here we used the Bayesian ideal observer model to generate predicted effects for the differences between the uniform and skewed distributions only and tested whether Acerbi et al.'s (2012) data would provide sufficient evidence for either the null or the alternative hypothesis. The researchers reported that the averages of the model fitted parameters for perceptual and motor noise were 0.157 and 0.072, respectively. Considering those estimates as the true parameter values, we examined what results we could expect in the actual experiment if we grossly over- and underestimate these values. In all simulations, there are 12 "participants", motor noise is defined as a scalar, and there is added measurement noise. For the measured effect for the difference in the standard deviation, we calculated the average difference between the uniform and skewed conditions in Acerbi et al.'s (2012) raw data – 16.69ms (SE=7.10 corrected for N=12).

We investigated the robustness of our analysis to under- and overestimation of the noise parameters. Underestimated parameter values were sampled from lognormal distributions centred at 1/5, 1/4, 1/3, and 1/2 of the true parameter values. Overestimated parameter values were sampled from lognormal distributions centred at 2 times greater, 3 times greater, 4 times greater, and 5 times greater than the true parameter values. To cover all possible combinations of parameter values listed above, as well as the possibility that only one of the two parameters is misspecified, we conducted a total of 80 simulations.

The results for the difference between the standard deviations of the uniform and skewed distributions provide evidence that this test is relatively robust to parameter misspecification. A total of 42 out of the 80 simulations provided moderate to strong evidence in favour of the alternative hypothesis (Table A1). There is a qualitative difference in the effects of motor and perceptual noise on the difference in standard deviations predicted by the model. Ignoring the effect of perceptual noise, as motor noise increases the difference in the variability of the responses in the two conditions decreases because responses move closer to veridical across both priors. In contrast, the effect of perceptual noise on the predicted difference in standard deviations follows an inverted U pattern. When perceptual noise is too low, the variability of responses across the two conditions is similar because responses are close to veridical, irrespective of the prior distribution. Conversely, when perceptual noise is too high, the variability of responses across the two conditions is again similar, but this time because responses are close to the mean of the prior distribution. Finally, when perceptual noise is neither too high, nor too low, the simulations provide evidence in favour of the alternative hypothesis. The simulations reached the sensitivity threshold of $BF_{10} > 3$ across all levels of

motor noise when perceptual noise was between 0.157 (true parameter value) and 0.471 (3×0.157). Looking at the other results, all simulations in which perceptual noise was 0.039 ($1/4$ of 0.157) and 0.031 ($1/5$ of 0.157) underestimated the expected difference in standard deviations and therefore provided inconclusive results. This was also the case for simulations in which there was a discrepancy between motor and perceptual noise – when perceptual noise was too high and motor noise too low, and vice versa. Nonetheless, the simulations suggest that we would still be able to find sufficient evidence in favour of the alternative hypothesis for a range of combination of noise parameter values. Furthermore, based on Acerbi et al.'s (2012) reported estimates of motor and perceptual noise, and the existing evidence for the relationship between physical and subjective time (Rammsayer & Pichelmann, 2018; Rammsayer, 2014; Lapid, Ulrich, & Rammsayer, 2008), we can be confident that arriving at such extreme values such as some of those used here will be very unlikely.

Table C1: Results for the difference in SD between the uniform and skewed conditions

(N=12). Columns one and two contain the mean of the sampling distribution of the perceptual and motor noise. Cells highlighted in blue signify simulations in which noise parameters were underestimated relative to their true value (shaded in white), and cells highlighted in red signify simulations in which the parameters were overestimated. The intensity of the colour signifies the absolute magnitude of deviation from the true parameters – the darker the colour, the greater the deviation. Column three contains the expected difference in the standard deviation of responses produced by the model, and column four contains the associated Bayes factor. In column four, cells highlighted with green signify simulations in which the results provide sufficient evidence in favour of the alternative hypothesis ($BF_{10} > 3$). Cells highlighted with orange signify simulations which resulted in $BF_{10} < 3$.

Mean of perceptual noise sampling distributions (SD)	Mean of motor noise sampling distributions (SD)	Simulated expected difference in SDs	BF_{10}
0.031 (0.002)	0.014 (0.001)	2.424	1.959
	0.018 (0.001)	2.315	1.900
	0.024 (0.001)	1.822	1.654
	0.036 (0.001)	1.252	1.412
	0.072 (0.001)	0.312	1.089
	0.144 (0.001)	0.245	1.069
	0.216 (0.001)	-0.056	1.017
	0.288 (0.001)	0.228	1.061
	0.360 (0.001)	0.015	1.002
0.039 (0.002)	0.014 (0.001)	3.813	2.839
	0.018 (0.001)	3.440	2.582
	0.024 (0.001)	3.108	2.366
	0.036 (0.001)	2.693	2.110
	0.072 (0.001)	1.166	1.378
	0.144 (0.001)	0.472	1.133
	0.216 (0.001)	0.178	1.051
	0.288 (0.001)	0.165	1.047
	0.360 (0.001)	0.114	1.029
	0.014 (0.001)	6.457	4.875
	0.018 (0.001)	6.094	4.596

0.052 (0.002)	0.024 (0.001)	5.627	4.226
	0.036 (0.001)	4.918	3.667
	0.072 (0.001)	2.744	2.140
	0.144 (0.001)	1.009	1.315
	0.216 (0.001)	0.300	1.082
	0.288 (0.001)	0.355	1.102
	0.360 (0.001)	0.259	1.069
0.0785(0.002)	0.014 (0.001)	10.766	7.371
	0.018 (0.001)	10.587	7.303
	0.024 (0.001)	10.528	7.285
	0.036 (0.001)	9.772	6.972
	0.072 (0.001)	6.863	5.184
	0.144 (0.001)	3.386	2.546
	0.216 (0.001)	1.635	1.573
	0.288 (0.001)	1.253	1.408
	0.360 (0.001)	0.783	1.235
0.157	0.014 (0.001)	15.549	8.043
	0.018 (0.001)	15.507	8.043
	0.024 (0.001)	15.602	8.042
	0.036 (0.001)	15.412	8.048
	0.144 (0.001)	11.568	7.611
	0.216 (0.001)	8.713	6.426
	0.288 (0.001)	6.676	5.043
	0.360 (0.001)	5.478	4.104
0.314 (0.002)	0.014 (0.001)	9.817	6.993
	0.018 (0.001)	9.899	7.030
	0.024 (0.001)	10.027	7.081
	0.036 (0.001)	10.112	7.121
	0.072 (0.001)	10.482	7.268
	0.144 (0.001)	11.126	7.484
	0.216 (0.001)	11.410	7.568
	0.288 (0.001)	11.049	7.460
	0.360 (0.001)	10.446	7.250
0.471 (0.002)	0.014 (0.001)	5.834	4.387
	0.018 (0.001)	5.625	4.221
	0.024 (0.001)	5.639	4.232
	0.036 (0.001)	5.805	4.367
	0.072 (0.001)	6.137	4.626
	0.144 (0.001)	7.385	5.565
	0.216 (0.001)	8.182	6.104
	0.288 (0.001)	8.717	6.424
	0.360 (0.001)	8.737	6.436
	0.014 (0.001)	3.686	2.749
	0.018 (0.001)	3.562	2.660
	0.024 (0.001)	3.471	2.599
	0.036 (0.001)	3.875	2.882

0.628 (0.002)	0.072 (0.001)	3.973	2.948
	0.144 (0.001)	4.722	3.510
	0.216 (0.001)	5.641	4.238
	0.288 (0.001)	6.338	4.783
	0.360 (0.001)	6.728	5.082
0.785 (0.002)	0.014 (0.001)	2.385	1.941
	0.018 (0.001)	2.343	1.915
	0.024 (0.001)	2.478	1.993
	0.036 (0.001)	2.438	1.966
	0.072 (0.001)	2.611	2.063
	0.144 (0.001)	3.319	2.497
	0.216 (0.001)	4.059	3.015
	0.288 (0.001)	4.667	3.472
	0.360 (0.001)	4.782	3.561

Appendix D. Instructions for tasks in Chapter 5, Experiment 1

1. Reproduction task

In this task, you will be presented with temporal intervals, and you will be asked to reproduce them. To start a trial, press the spacebar. A white fixation cross is going to appear in the middle of the screen, keep looking at the cross for the entire trial. Shortly after that, a cloud of black and white dots will be presented on the screen for a variable amount of time. The position and configuration of the dots on the screen will change between trials, but this is not related to the task you have to complete. Shortly after the dots disappear, the fixation cross will turn green. This is your cue that you can begin your response. To reproduce the interval, press and keep the spacebar on the keyboard pressed for as long as you think the dots were presented. It is important that you do not try to explicitly count how long the stimulus was presented for. All durations in the experiment are very short and difficult to count. Just pay attention to how long the dots remain on the screen for. Some time after you provide a response, you will receive feedback [here participants were shown the feedback display]. The coloured line represents how your response was from the true duration of the interval. If the coloured line is exactly over the white line in the middle of the box, this means there was no error between your response and the presented interval. If the coloured line is to the left of the white line, it means your response was shorter than the presented interval. If the colour line is to the right of the white line, it means your response was longer than the presented interval. The further the coloured line is from the middle, the bigger the difference between your response and the stimulus. The colour of the coloured line also provides you with some information – if the line is green, it means your response was within 10% of the presented duration. If the line is brown, it means your response was further away. Your goal is to try to be as accurate as possible.

There are a few instances in which trials will be excluded and repeated later in the block. If you start reproducing the interval before the response cue appears, the trial will be excluded, and a message saying “Too soon” will appear on the screen. If you start reproducing the interval longer than 1.5s after the response cue appears, the trial will be excluded, and a message saying “Too late” will appear on the screen.

The task will start with a training block consisting of 35 trials. In the training block, you will be presented with the stimuli in ascending order (from shortest to longest). After that, you will complete between 4 and 8 testing blocks depending on your performance. Each testing block will consist of 50 trials. It will contain the same stimuli as in the training block, but their order will be shuffled. At the end of each block, the script will calculate the correlation between the presented durations and your responses. If you manage to complete 5 blocks in which you exceed the needed threshold, you will have a 15-minute break, after which you will do the second part of the duration reproduction task. The task will be the same, but the durations will be different. If you do not manage to reach the needed threshold in any 4 blocks, this will be the end of the experiment.

The task is difficult and demanding. You can take as many breaks as often as you like.

2. Discrimination task

In this task, you will be presented with two temporal intervals one after the other, and you will be asked to report which interval was shorter or longer in different blocks. When you are ready to start the trial, press the spacebar. A fixation cross will appear shortly after, keep looking at the fixation cross throughout the whole trial. Then, a cloud of black and white dots will be presented for some time. This will be the first interval. The position and configuration of the dots do not matter. After the first interval ends, the screen will show the fixation cross alone for some time. Then, the cloud of dots will appear again for a different duration. This will be the second interval. It is important to not use any strategies to time the duration of either interval. After the second interval ends, you will be asked to report which stimulus was shorter or longer – the question on the screen will specify which, make sure you read it carefully. If you think the correct answer is interval 1, press “1”. If you think the correct answer is interval 2, press “2”. At this point, you can change your response. When you are happy with your choice, press the spacebar to save your response.

You will complete 2 blocks, each will be 70 trials long. This is expected to take ~20 minutes. You can take a break at any point.

3. EEG task

This task has the same structure as the duration reproduction task you completed earlier but you will not be required to reproduce the durations of the stimuli. Every trial will begin with the presentation of a white fixation cross. Keep looking at the fixation cross throughout the entire trial. After some time, a cloud of black and white dots will be presented on the screen for some time. You will not be asked to reproduce the duration of the stimulus, but still pay attention to the presented duration as if you would have to provide a response. Do not explicitly count how long the stimulus is presented for. On some trials, the fixation cross will be presented in green instead of white. Your task is to detect those changes as fast as possible. When you see the fixation cross changes colour, press the spacebar.

The session will consist of 4 sections. Each section will consist of 6 blocks. In the first block, you will see the durations from the current condition in ascending order, so that you get used to the distribution. In the rest of the blocks, the stimuli will be shuffled. Each block will take around 3 minutes to complete. There will be two scheduled 15-minute breaks at the end of section 1 and section 3. You can take additional breaks at the end of each block. During trials, please keep your face relaxed, do not clench your teeth, and try to avoid rocking your feet or fidgeting – these will affect the EEG signal. Finally, blink naturally, do not try to suppress blinks.

Appendix E. Distribution of regression to the mean in Chapter 5,
Experiments 1 and 2

Figure E1: Distribution of regression to the mean Chapter 5, Experiment 1. Regression to the mean is calculated as the difference in responses to the 626ms stimulus when it is presented in the context of long compared to short durations. The histogram shows the number of participants falling within each regression to the mean bin, with the data divided into bins centred at -15 to 135ms in steps of 10ms.

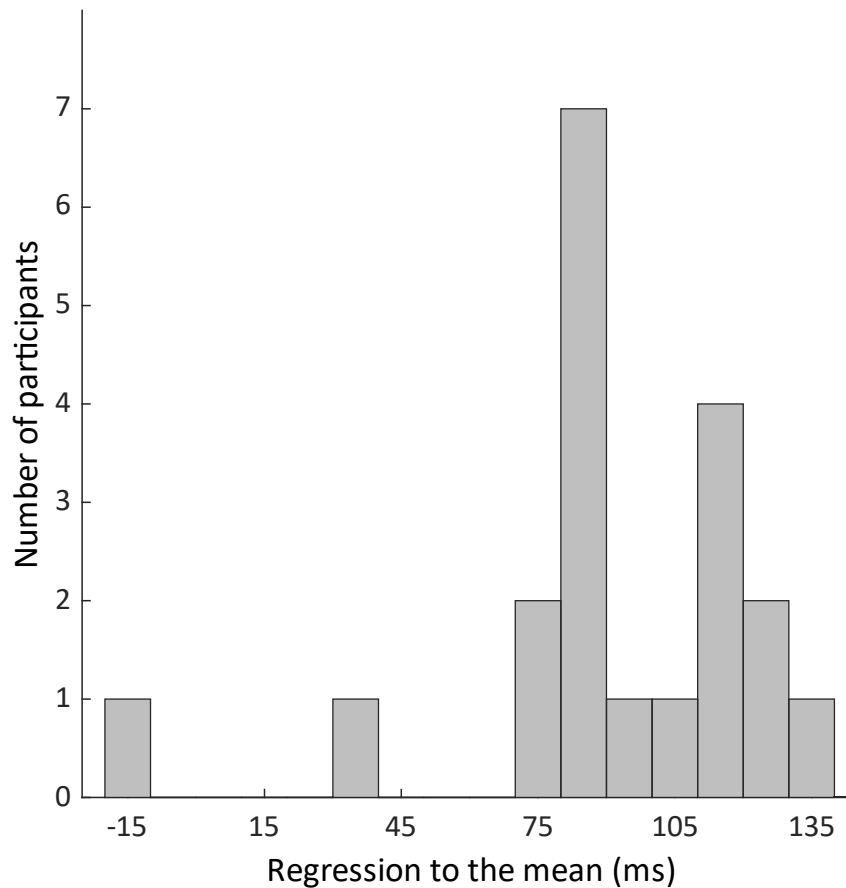
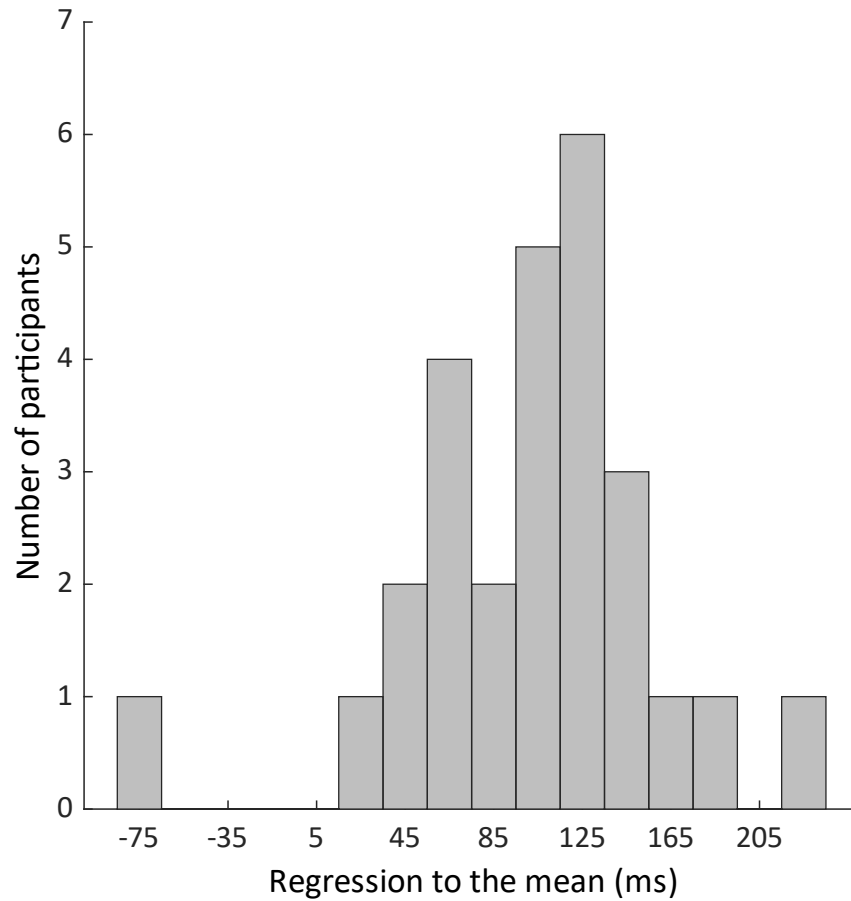
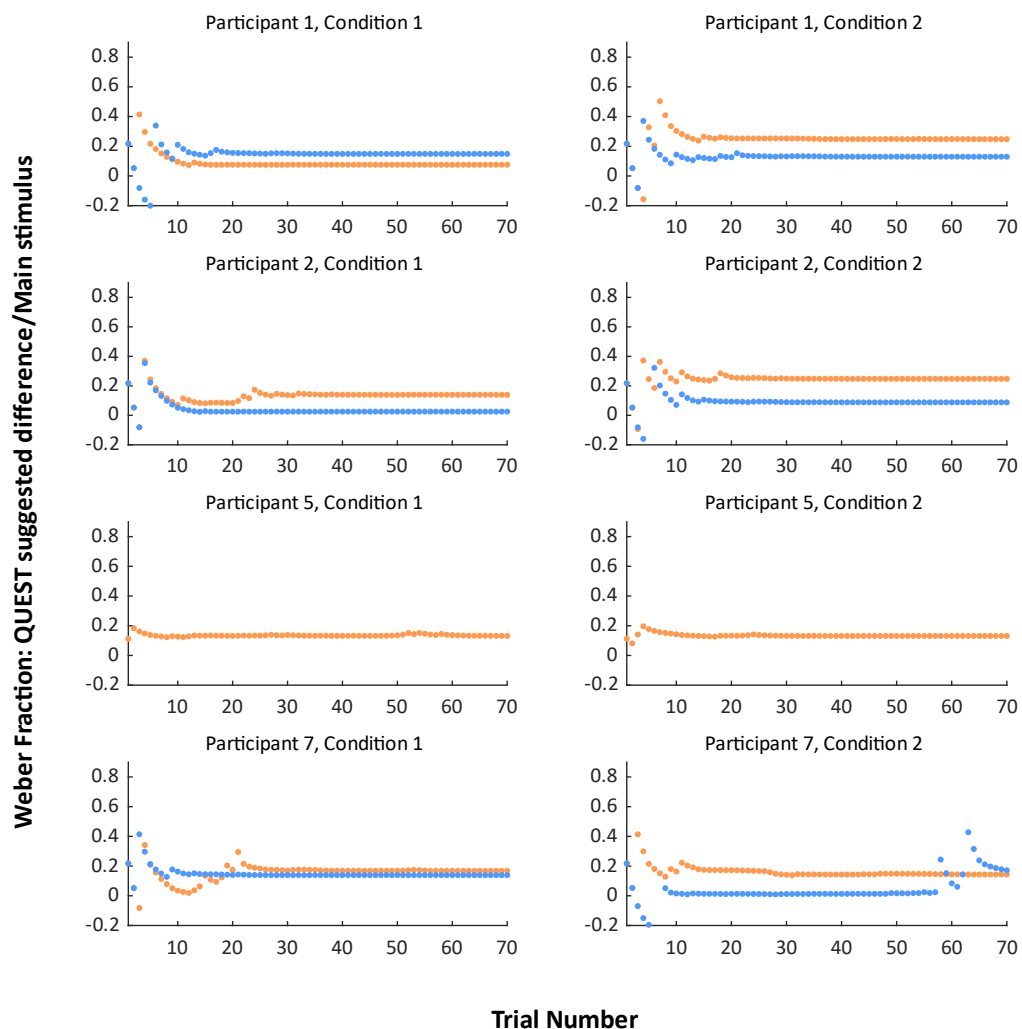


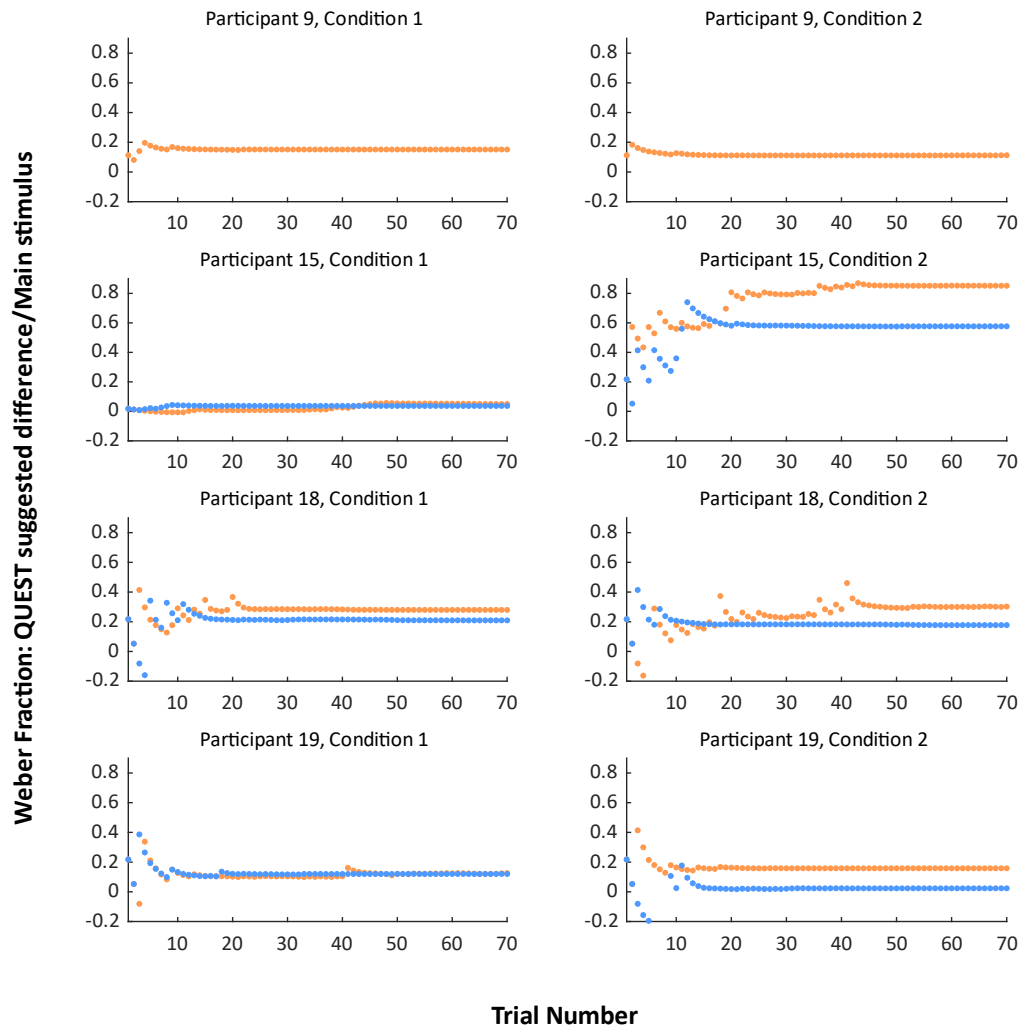
Figure E2: Distribution of regression to the mean Chapter 5, Experiment 2. Regression to the mean is calculated as the difference in responses to the 626ms stimulus when it is presented in the context of long compared to short durations. The histogram shows the number of participants falling within each regression to the mean bin, with the data divided into bins centred at -75ms to 225ms in steps of 20ms.

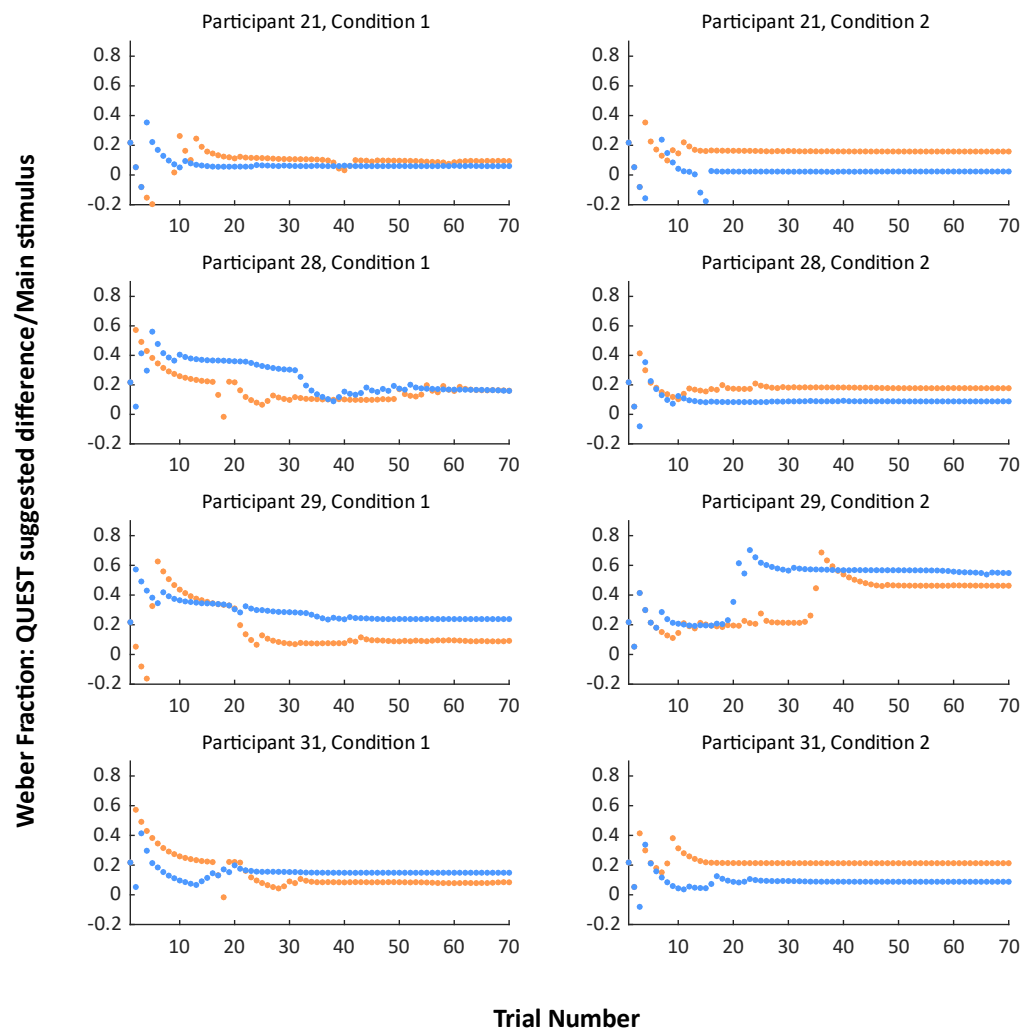


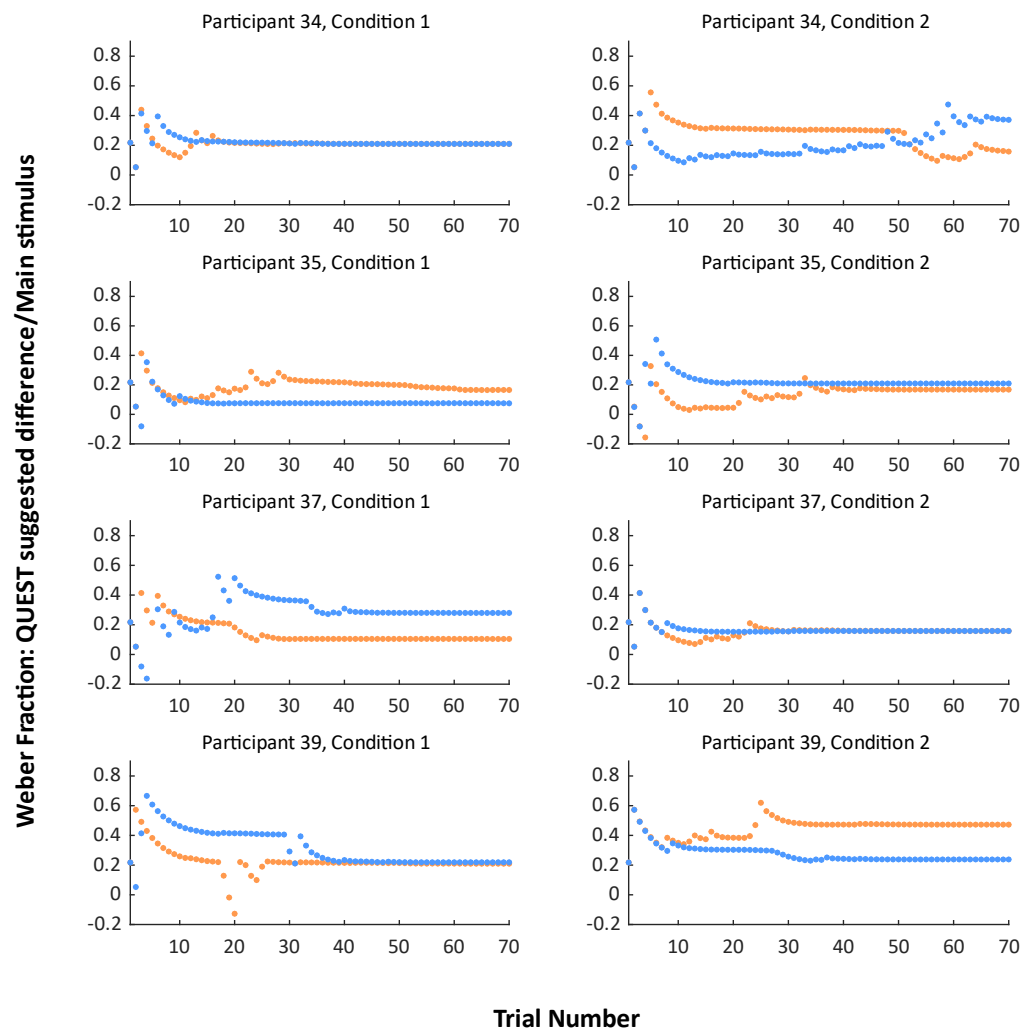
Appendix F. 2IFC threshold convergence and distribution Chapter 5, Experiment 1

Figure F1: Trial-by-trial Weber fractions suggested by QUEST individually for each participant and block across the two conditions in Chapter 5, Experiment 1. On each trial of the 2IFC task, QUEST provided a suggestion for the difference between the main stimulus (626ms) and the comparison stimulus that should be tested next based on the participant's performance on the previous trials. Sometimes QUEST would suggest a difference smaller than 0 in which case the difference suggested by QUEST did not correspond to the actual difference in that trial (which was drawn from a uniform distribution between 6 and 18ms). Participants completed 2 blocks per condition. Trials from the first block are depicted in orange, trials from the second block are depicted in blue. In Condition 1 participants were asked to identify the longer of the two stimuli presented on each trial, in Condition 2 – the shorter of the two stimuli presented on each trial.









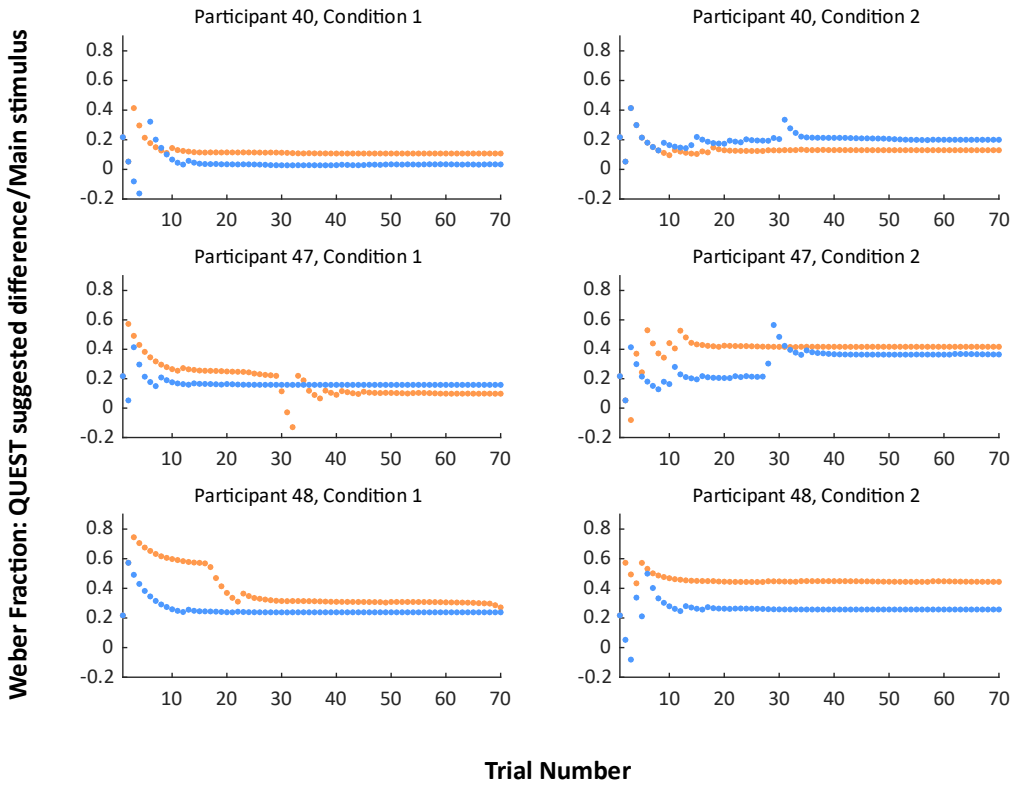


Figure F2: Distribution of Weber fractions based on performance on the 2IFC task in Chapter 5, Experiment 1, calculated following the pre-registered analysis plan. Weber fractions are calculated by dividing the average 2IFC threshold of each participant by the main stimulus in the discrimination task – 626ms. The histogram shows the number of participants falling within each Weber fraction bin, with the data divided into bins centred at 0.095 to 0.38 in steps of 0.015.

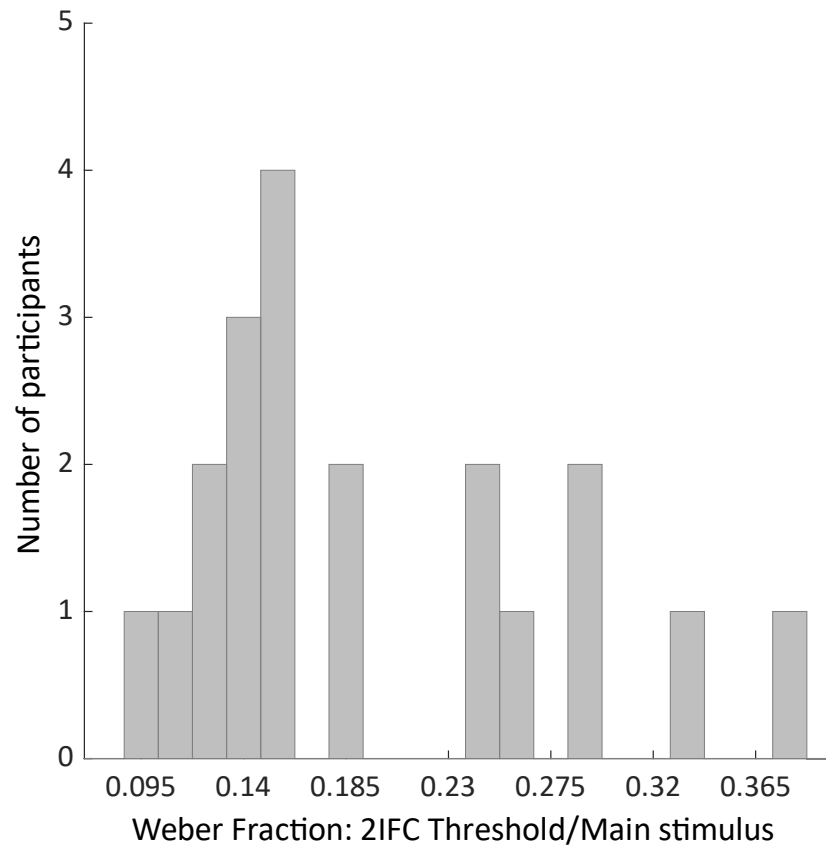
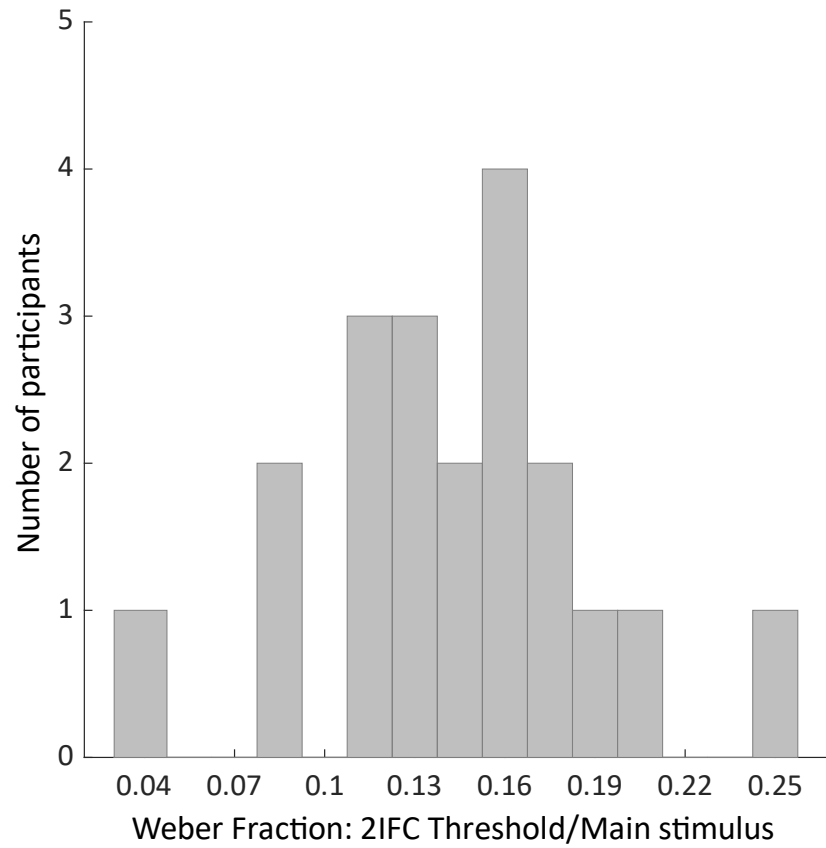


Figure F3. Distribution of Weber fractions based on performance on the 2IFC task in Chapter 5, Experiment 1, calculated following the exploratory analysis plan. Weber fractions are calculated by dividing the average 2IFC threshold of each participant by the main stimulus in the discrimination task – 626ms. The histogram shows the number of participants falling within each Weber fraction bin, with the data divided into bins centred at 0.04 to 0.25 in steps of 0.015.



Appendix G. Instructions for tasks in Chapter 5, Experiment 2

1. Estimation task

[Start of task]

In this task, you will be asked to estimate how long a visual stimulus stays on the screen for. Please sit 60cm away from your monitor.

The stimulus is an array of black and white dots. On every trial, the dots will change position on your screen. Their position does not matter - it is not relevant to your task. Your task is to estimate how long the dots are on the screen. To start a trial, you will have to press Space. Each trial starts with the presentation of a green fixation cross. During the trial, keep looking at the cross. Press Space for a demonstration of a trial.

[Participants are shown an example trial. The cloud of black and white dots is presented, followed by the scale used to record responses.]

To provide your estimate of the duration of the dots, click on the scale and drag the marker to the left or to the right until you are happy with your selection. When you have selected your response, click on the number below the scale to save it. Try it out below. (In the experiment you will have 5 seconds to provide a response.)

[Participants provide a response and are shown different examples of the feedback display.]

After providing your response, you will receive graphic feedback showing how accurate you were on that trial. The white line in the middle of the box represents perfect accuracy. The coloured line represents your accuracy. The further away the coloured line is from the white midline, the bigger the difference between your response and the true duration of the stimulus.

Example 1: If the coloured line is to the right of the midline, your response was longer than the stimulus.

Example 2: If the coloured line is to the left of the midline, your response was shorter than the stimulus.

The colour of the feedback line also provides information about the accuracy of your response. If the feedback line is red, your response was very far away from the duration of the stimulus. If the feedback line is green, your response was close to the duration of the stimulus.

Example 1: Your response was very far from the true stimulus duration.

Example 2: Your response was close to the true stimulus duration.

It is important that you do not try to count how long the dots are on the screen for. All durations you will see will be very short and very hard to count, but please do not try to use any counting strategies to complete the task. Just pay attention to how long the dots remain on the screen.

The experiment is divided into 4 blocks. Training blocks: Blocks 1 and 3 are for training so that you get used to the task. Each training block will take about 5 minutes to complete. Testing blocks: Blocks 2 and 4 are the testing blocks. Each testing block will take about 20 minutes to complete. Performance: At the end of each block, the script will calculate the correlation between the presented durations and your responses. If the correlation at the end of a block is below a certain threshold, the experiment will end, and you will not complete the rest of the experiment.

Breaks: There will be multiple fixed breaks throughout the experiment. You cannot skip these breaks, you cannot take longer than the allocated time, and you cannot take breaks at any other times. A counter will display the remaining break time. An alarm will sound to signal the end of a break, but do not rely solely on this. Make sure you are at your computer when the break ends because the experiment will restart.

Missing trials: The maximum allowed "break" between trials will be 15 seconds. If you do not start a trial within this period, the trial will be marked as missed. If you fail to provide a response 5 seconds after the response screen appears, the trial will be marked as missed as well. If you miss more than 10 trials, the experiment will end. Do not worry about these time restrictions - within this experiment, 5 seconds is a very long time.

Computer errors: Trials will also be discarded if the computer fails to present stimulus durations accurately. If the computer fails on more than 20 trials, the experiment will end. You will be reminded of these instructions at the start of each block.

[Start of blocks 1 and 3 - training block]

This block is for training. It consists of 40 trials and will take about 5 minutes to complete.

[Start of block 2 – first testing block]

This is a testing block. It consists of 250 trials and will take about 20 minutes to complete.

Breaks: There will be a 2 min break in the middle of the block and a 15 min break at the end of the block. You cannot skip the breaks and you have to be ready to continue when the breaks end. A counter will display the remaining break time. An alarm will sound to signal the end of the break, but do not rely solely on this. Make sure you are at your computer when each break ends because the experiment will restart.

Skipping a trial: In testing blocks, on each trial, you are allowed to skip INSTEAD of providing a response. To skip a trial, press S when the response screen appears. Skip trials if you did not pay attention to the intervals. Skipped trials will be presented later on. You are allowed to skip a maximum of 10 trials in each block.

[Start of block 4 – second testing block, last block in the estimation task]

This is the last testing block. It consists of 250 trials and will take about 20 minutes to complete

Breaks: There will be a 2 min break in the middle of the block and a 5 min break at the end of the task. After this break, please start the discrimination task (the other file in the folder). You cannot skip the breaks and you have to be ready to continue when the breaks end. A counter will display the remaining break time. An alarm will sound to signal the end of the break, but do not rely solely on this. Make sure you are at your computer when each break ends because the experiment will restart.

2. Discrimination task

[Start of task]

In this task, you will be asked to compare two temporal intervals. On every trial, you will be shown two intervals one after the other. Then, you will be asked which interval was SHORTER [or LONGER in the other condition]. Altogether you will do between 3 and 5 blocks depending on your performance. Each block takes 10 minutes to complete. Please sit 60cm away from your monitor.

'During the experiment, a green fixation cross will be presented in the middle of the monitor. Please, keep looking at the cross throughout the entire trial. A prompt will appear to indicate when it is time to begin a trial. To start a trial, press Space. Pressing Space before the prompt appears will have no effect. After you press Space, an array of black and white dots will appear on the screen for some time. When the array disappears, there will be a brief pause after which a second array of black and white dots will appear on the screen. Press Space for a demonstration of the stimulus.

[Participants are shown an example trial. A cloud of black and white dots is twice, one after the other, followed by the response screen.]

You will be asked to compare the two intervals. To provide your response, click on the scale and select 1 for Interval 1 or 2 for Interval 2. When you have selected your response, click on the button below the scale to save it. (In the experiment you will have 5 seconds to provide a response.)

[Participants provide a response and are shown the feedback display.]

After providing your response, you will receive feedback on whether your response was correct. If your response was correct, the feedback message will say 'Correct' in green. If your response was incorrect, the feedback message will say 'Incorrect' in red.

Make sure that when you start the trial, you are ready and paying full attention. If on any trial you feel like you did not pay attention to the stimulus, instead of responding you can skip the trial by pressing S. Skipped trials will be repeated later on. You can skip a maximum of 10 trials in each block. It is important that you do not try to count how long the dots are on the screen for. All durations you

will see will be very short and very hard to count, but please do not try to use any counting strategies to complete the task. Just pay attention to how long the dots remain on the screen.

[Start of training block]

In this condition, you will be asked which of the two intervals is SHORTER [or LONGER in the other condition]. This block is for training and consists of 16 trials.

[Start of testing block]

In this condition, you will be asked which of the two intervals is SHORTER [or LONGER in the other condition]. Each block contains 90 trials and will take about 10 minutes to complete.

Skipping a trial: On each trial, you are allowed to skip INSTEAD of providing a response. To skip a trial, press S when the response screen appears. Skip trials only if you did not pay attention to the intervals. Skipped trials will be presented later on.\n You are allowed to skip a maximum of 10 trials in each block.

Missing a trial: If you do not start a trial within 15 seconds of the start prompt, or you do not respond within 5 seconds of the response prompt, the trial will be marked as missed. If you miss more than 10 trials, the experiment will end.

Breaks: There will a 2 min break at the end of each testing block. You cannot skip the break and you have to be back when the break ends. A counter will display the remaining break time. An alarm will sound to signal the end of the break, but do not rely solely on this. Make sure you are at your computer when the break ends because the experiment will restart.

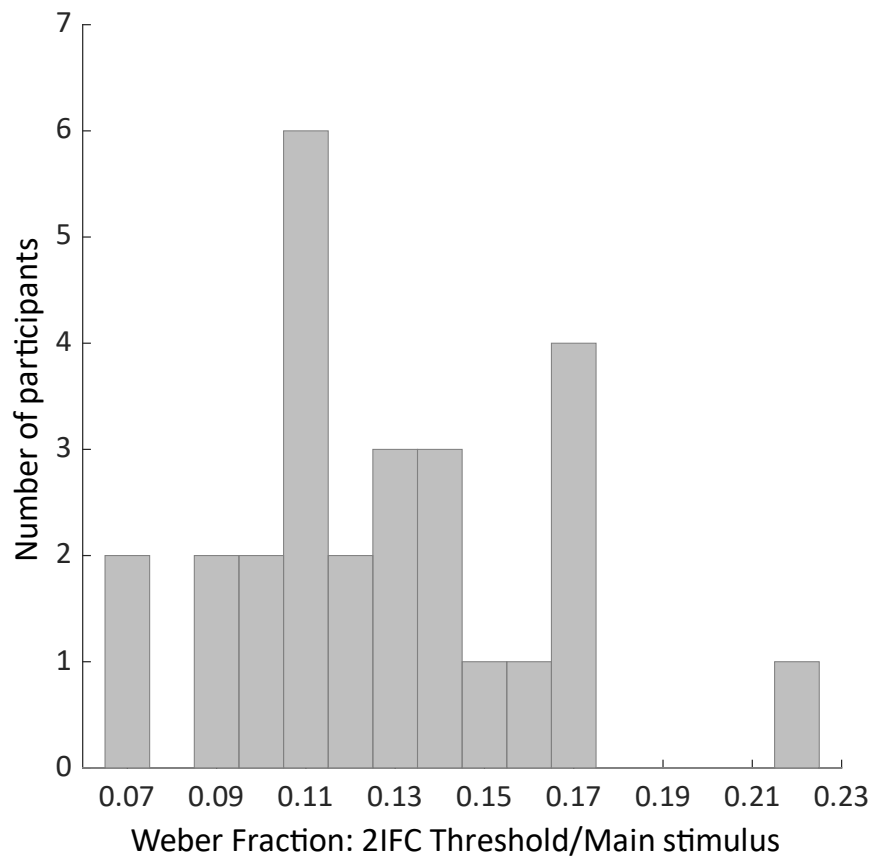
Computer error: Sometimes the computer might not display the desired durations accurately. If this happens more than 20 times, the experiment will end.

Please do not use any counting strategies to estimate how long the stimuli are presented for!

Please sit 60cm away from your monitor.

Appendix H. 2IFC threshold distribution Chapter 5, Experiment 2

Figure H1. Distribution of Weber fractions based on performance on the 2IFC task in Chapter 5, Experiment 2. Weber fractions are calculated by dividing the average 2IFC threshold of each participant by the main stimulus in the discrimination task – 626ms. The histogram shows the number of participants falling within each Weber fraction bin, with the data divided into bins centred at 0.07 to 0.23 in steps of 0.01.



Appendix I. Instructions for tasks in Chapter 6

1. Contrast discrimination

In this task, you will be asked to detect faint stimuli presented on the screen. You will have to press space to start each trial. On each trial, you will be presented with two intervals, one after the other. In both intervals, a black and white checkerboard will be presented on the screen. During one of the two intervals, a black and white grating will be presented in the same place as the checkerboard. After the second interval ends, your task will be to report which of the two intervals contained the grating. Press “1” for interval 1, or “2” for interval 2. At this point, you can change your response. Once you are happy with your choice, press space to save your response.

[Session 1] You will complete 20 blocks.

[Sessions 2 to 5] You will complete 10 blocks.

Each block will contain 20 trials, and it will take around 3 minutes. Please do not move your head from the chinrest during the block because this will affect the calibration. You will have the opportunity to move at the end of each block. If you get tired during a block, try to close your eyes at the end of a trial.

2. Duration estimation training

In this task, you will be asked to estimate the duration of a black and white grating presented on the screen. In each trial, you will be presented with a black and white grating that will stay on the screen for a variable amount of time. Do not explicitly count or time how long the grating is presented for, just pay attention to how much time passes while the grating is on the screen. After the grating disappears, you will be asked to report how long the grating was presented for. A randomly selected duration will appear on the screen, on every trial the duration will be different. To change the suggested duration, move the mouse to the left or to the right. Moving the mouse to the left will decrease the duration, moving the mouse to the right will increase the duration. When you are happy that the number displayed on the screen reflects the duration of the grating presented in the trial, press space to save your response. You will then be given feedback showing the actual duration of the grating. Press space to start the next trial.

[Session 1] You will complete 10 blocks.

[Sessions 2 to 5] You will complete between 3 and 10 blocks.

Each block will contain 21 trials and will take around 4 minutes. Please do not move your head from the chinrest during the block because this will affect the calibration. You will have the opportunity to move at the end of each block. If you get tired during a block, try to close your eyes at the end of a trial.

3. Main duration estimation task

In this task, you will be asked to estimate the duration of a black and white grating presented on the screen. The task will be very similar to the duration estimation

training task. On every trial, you will be presented with a black and white grating. Do not explicitly count or time how long the grating is presented for. On some trials, a black and white checkerboard will be presented in addition to the grating, and you may not be able to see the grating.

When the stimuli disappear, you will be asked to report how long the grating was on the screen. A randomly selected duration will appear on the screen, on every trial the duration will be different. To change the suggested duration, move the mouse to the left or to the right. Moving the mouse to the left will decrease the duration, moving the mouse to the right will increase the duration. When you are happy that the number displayed on the screen reflects the duration of the grating presented in the trial, press space to save your response. On trials where the black and white checkerboard was presented and you did not see the grating, you still have to provide a response for the duration of the grating. In these situations, you will have to provide your best guess of what the duration of the grating might have been. It is important to note that the duration for which the checkerboard is presented is not indicative of the duration of the grating, so do not use the duration of the checkerboard to estimate the duration of the grating. In this task, you will not receive feedback regarding the accuracy of your estimate.

After you provide your duration estimate, you will be asked to report how clear your experience of the black and white grating was. There will be four available responses: 1 – no experience, 2 – weak experience, 3 – almost clear experience, and 4 – clear experience. “4” would correspond to a clear, non-ambiguous experience of the grating being presented. “3” would correspond to having a weaker experience of the stimulus than “4” – an ambiguous experience of the stimulus, with some aspects of the stimulus being experienced more vividly than others. “2” would correspond to feeling that something has been shown, but without being able to characterize any of its features further. “1” would correspond to not seeing the grating at all.

You will complete 7 blocks. Each block will contain 63 trials, and it will take around 10 minutes to complete. Please do not move your head from the chinrest during the block because this will affect the calibration. You will have the opportunity to move at the end of each block. If you get tired during a block, try to close your eyes at the end of a trial.

Appendix J. Instructions for tasks in Chapter 7

You will complete 3 blocks of a time perception task, in which you will be presented with visual stimuli with variable durations. Each block will start with 20 trials for training. In the training, you will be presented with 10 repetitions of a short interval, and 10 repetitions of a long interval. Your task will be to learn to tell apart the short from the long intervals. On each trial, a flash will appear on the screen. After the flash disappears, you will be asked to report whether the stimulus was the long or the short stimulus. You will use the buttons J and N to provide your responses – which button corresponds to “short” and which button corresponds to “long” will be displayed on the screen. After you provide your response, you will receive feedback regarding the accuracy of your response. A message saying “Correct” will be displayed if you provide the correct response, and a message saying “Incorrect” will be displayed if you provide the wrong response.

In the testing part of the block, you will be presented with a range of durations between the short and long durations you saw in the training block. Again, on every trial, you will be presented with a flash. After the flash disappears, you will be asked to report whether the stimulus was more similar to the shortest or to the longest interval you saw in the training. You will not be provided with feedback in the test trials.

It is essential that you do not use any techniques to explicitly count the stimuli.

Do not move your head from the chinrest during a block because this will interfere with the eye-tracking and we will have to start the block again. If your eyes get tired, you can have a break by closing your eyes. You will have the opportunity to move at the end of each block. Each block will take ~20 minutes to complete.