LEARNING PRIORS FOR BAYESIAN PERCEPTION IN NORMAL AND ATYPICAL DEVELOPMENT

Lucy Anne McKeown (MSci, MSc)

Abstract

Over recent years, biases of sensory magnitude estimations, such as central tendency bias have been thought to reveal the weighting of a Bayesian prior as estimations drifted towards the mean of the stimulus distribution.

In temporal estimation specifically, Vierordt's law described the underestimation of long durations and the over estimation of short durations. Vierordt's law has since been interpreted via a Bayesian explanation by Jazayeri and Shadlen (2010). This suggested that statistical regularities were learned and incorporated into a prior, but it did not provide insight into how priors develop over time.

Priors may update through the dynamic estimation of the mean of the stimulus distribution. Or, as suggested more recently, that the prior is updated according to the value of the recent stimulus using a Kalman filter. Glasauer and Shi (2019; 2021b) demonstrated the possible use of the Kalman filter in participants because central tendency effects were reduced for the random walk (gradually changing magnitude) stimuli relative to fully random stimuli. However, the current research suggested that, during a temporal reproduction task, models that updated using the mean of the stimulus distribution provided a much better fit to the data than a simple iterative (Kalman filter) model.

The extent of the central tendency bias in temporal reproduction was also correlated with AQ trait scores and showed a relationship between higher AQ trait scores and lower compression for random walk sequences. Therefore, higher levels of autistic traits may be associated with differences in prior acquisition but a tendency for a static or iterative model according to the number of autistic traits was not found.

Temporal perception is necessary for learning when sensory stimuli appear and how long they persist, in addition to the functional benefits of multisensory integration. However, the possibility that spatial quantities may be more likely to induce iterative strategies than temporal quantities was exemplified by research into path integration (Glasauer & Shi, 2021b; Petzschner & Glasauer, 2011) and visually guided reaching (Verstynen & Sabes, 2011). Using spatial localisation (the prediction of target locations

from a cue location), the current research was the first demonstration of flexibility in prior acquisition strategy in line with stimulus features. This was only found in adult participants however, as older children did not show this same level of flexibility and the younger children displayed a type of flexibility that may be a reactive response to the stimuli.

The exploration of bias in a single stimulus has been the focus of most research using the Bayesian observer model. However, some Bayesian model research used the time order error, as a metric for understanding different priors in autistic individuals (Sapey Triomphe et al., 2021). The time order error (TOE) occurs when the first stimulus is perceived with relatively more bias (over and underestimation) than the second stimulus. In the current research, a Bayesian model accounted for a TOE outside the context of a two-stimulus comparison, which was not explained by previous models and may imply that passive memory processes provide a sufficient explanation without the need for processes rooted in interference.

Taken together, these results show that Bayesian models provided a relatively good fit for a range of different prior acquisition strategies across different temporal and spatial tasks. Although the current research found a preference for static models in temporal research and some preference for a previous trial model in spatial research. A degree of flexibility, not yet seen in the literature, was displayed in the spatial task regarding participants prior updating strategy.

Acknowledgements

The completion of my thesis has been a challenging but enriching process. I am very grateful to many, many people.

I would like to thank my supervisor, Neil Roach for an incredibly valuable and formative learning experience**.** Through our meetings, I felt that I left truly understanding the concepts we discussed and ready to continue with work that had previously felt overwhelming. From your example, I learned patience, resilience, and thorough analysis. I would additionally like to thank my second supervisor Dr Danielle Ropar. Her encouragement and expertise were essential in the progression of my research skills and confidence.

I would also like to thank the University of Nottingham Psychology department for creating an environment that I would enjoy working in for a total of 8 years. I will look back on my time there with fondness. The halls of the building were overflowing with warmth from the staff and student occupants. A special thank you to the MGS Economic and Social Research Council and their team in Nottingham for supporting the project.

My family fiercely supported my desire to study a PhD in psychology. I cannot thank my parents enough for encouraging me to follow my academic dreams and for offering me their unwavering kindness. They gave me the confidence to believe in my personal strengths and held space for me to choose my own path in life. I would also like to individually thank my brother, Will, for our research discussions. He commenced his PhD study 2 years prior and although working in vastly different fields, the similarities in our work propelled some intriguing debates.

Another important mentor throughout this time was my sport psychologist Rebecca Chiddley. She taught me many stress management strategies and how to implement them seamlessly into my life. As a result, continuing competitive dressage and my post graduate studies was possible and more importantly, she helped me manage my expectations when my sport had to take a back seat.

A special thanks must be given to my friends in Nottingham and back home in Stafford. To the unfortunate souls who asked what my PhD was about, I hope the possible excitement or frustration in my voice (depending on when you asked me) was not too disconcerting.

Finally, thank you to my partner Sukhraj. Thank you for the confidence you instilled in me, even when I lost sight of it myself. I am incredibly grateful for how you encouraged me to keep moving forwards, especially when the perfectionism set in. On those difficult days where we were both facing inevitable PhD stressors, you provided consistent level headedness, support, and encouragement. The process of two people working on a doctorate in the same house can be intense, but I greatly enjoyed and appreciated those times.

Thank you to you all.

Dedication

I would like to dedicate this thesis to my late grandfather, Mr Charles Henry Jones. His curious and generous nature led him to support the passions of many people, myself included.

Covid impact statement.

The most significant impact of the COVID-19 pandemic on this research was the requirement to move experiments online. Whilst this was an incredibly useful skill to learn, it slowed my progress significantly because I would usually benefit from the support of my supervisor in the experiment programming software that we were using in the lab pre-COVID (MATLAB). Learning to programme experiments does, by its very nature involve a lot of self-directed learning, however with the support of peers and my supervisor I would have been able to prevent and or solve coding obstacles much more quickly than I was able to alone. During this transition period, I had a lot to learn about programming an experiment in a relatively short space of time. Online courses and forums have been my main source of learning about Psychopy Palovia but receiving support for specific coding issues could take up to days or weeks, simply due to how busy the online forums were becoming during this time.

Contents

List of Tables

List of Figures

1. ***Literature review***

1.1 Temporal estimations: properties

Time perception provides humans the ability to sense the initiation, duration and termination of sensory signals. This means that estimates can be made according to when an event or time interval might occur or how long it may last. As a result, humans have the ability to navigate through the world and participate in activities such as sports, music, dance. Time perception also has a specific role in processing and integrating signals from multiple sensory modalities (Vroomen & Keetels, 2010). Humans rely on multi-sensory integration for interacting with objects in space and judging the movement of these objects in relation to themselves (Matthews & Meck, 2016). The perception of time intervals is therefore crucial for many daily activities where multiple sources of sensory information must be integrated, for example when crossing the road or comprehending speech (Burr & Morrone, 2006).

Overall, implying that timing is crucial for perceiving and moving through a coherent world. Due to its importance, researchers have used psychophysical methods to understand the way in which the brain represents time.

Scalar property

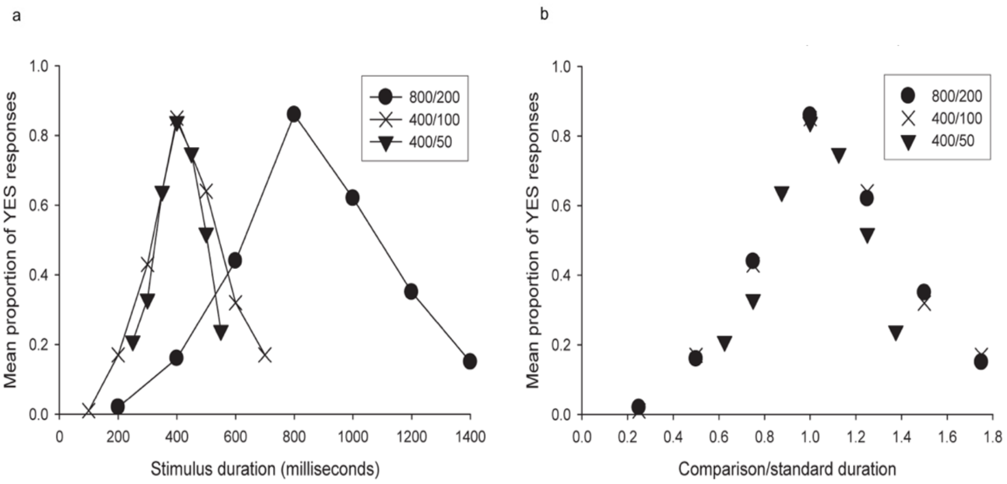
The relationship between the physical magnitude of a time interval and the perceived quality of that temporal stimulus has been described as scalar (Allman, Teki, Griffiths, & Meck, 2014; Church, 1984; Gibbon, Church, 1984). There are two key properties of scalar timing as defined by Wearden and Lejune (2008) and Wearden (2003). The first property of mean accuracy refers to the first feature of estimates that conform to scalar timing that is, estimates are correct on average. This is shown in a temporal generalisation task where participants are more likely to be correct in estimates of equality when the standard and the comparison become increasingly similar in value (see Figure 1a). The second property of duration estimates that conform to scalar timing is referred to as the scalar property of variance. This is produced from Weber's law because this property requires that as the physical stimulus increases, there is a proportional increase in the standard deviation of the estimate (Wearden, 2003). As a result, the coefficient of variation (SD/M) can be plotted as linear to demonstrate this property (Wearden & Lejeune, 2008).

The presence of the scalar property is exemplified by the process of the superposition (Wearden & Lejeune, 2008). The superposition of temporal data can be seen once two different data sets are collected and plotted against relative time**.** Participants experienced a temporal generalisation, a task where a standard value is presented before a comparison value, and they were asked to judge whether it was the same value as the standard. This allowed for a temporal generalisation curve to be plotted that is, the stimulus duration of the comparison plotted as a function of the proportion of 'yes' responses. The property of scalar timing predicts that the proportion of 'yes' responses will increase until the comparison duration is equal to the standard and then decrease again at longer durations. The result is a peak in this curve at the standard duration (See Figure 1a). The process of superposition is revealed when sets of durations with different standards are plotted together on a figure where the x axis is relative time (comparison divided by the standard), but the y axis remains as the proportion of 'yes' responses. In the example shown here (see Figure 1b) they used one set of durations with a standard of 400 and steps of 100, another set of durations with a standard of 400 and steps of 50 and a third set of durations with a standard of 800 and steps of 200.The scalar properties of timing are demonstrated by superposition in this figure because the peaks align when plotted against relative time, despite different sets of physical durations being used. Although the research described here pertains to animal studies, there are some studies evidencing the scalar properties of timing in research with children (Droit-Volet, Clément, & Wearden, 2001) and with adults (Wearden, 1991; Wearden & Lejeune, 2008)**.**

Violations of the scalar property of timing have been demonstrated when pharmacological dopaminergic interventions led to altered timing (Buhusi & Meck, 2002; Meck, 1996). Other potential extraneous sources of non-scalar noise arise from the motor noise present in responses (Fulcher, 2017). This type of noise has a larger influence at short durations but it is negligible compared to other sources of variability at longer durations, therefore it may interfere with the scalar property (See supplementary information in Roach, McGraw, Whitaker, & Heron, 2017).

Some violations are interesting in themselves because they signal stable individual differences in time perception. Individuals with Parkinson's disease, show systematic overestimation of short durations and underestimation of long durations (Malapani et al., 1998). Furthermore, developmental disorders such as autism, have been associated with differences in weber ratio and quality of superimposition in bisection performance (Allman, 2011). Additionally, differences in this property have been demonstrated in major affective disorder and schizophrenia (Penney, Meck, Roberts, Gibbon, & Erlenmeyer-Kimling, 2005).

Figure 1. The principle of superposition.



Note. Reprinted from "Scalar Properties in Human Timing: Conformity and Violations", by J.H. Wearden and H. Lejune 2008., *Quarterly journal of experimental psychology, 61* (4), pg 3. Copyright 2008 by the Experimental Psychology Society.

a) Mean proportion of 'yes' responses plotted as a function of absolute stimulus duration. The three duration sets shown here were purposely different in terms of the standard used and or the increments between the intervals to show how superposition occurred during peak interval procedures.

The superposition process is demonstrated by b) the mean proportion of yes responses plotted as a function of relative time (comparison/standard). The peak of all duration sets now superpose.

Modality effects

Psychophysical methods have been used to understand explanations as to why human perceptions of time are not identical to their physical magnitude. For example, the ventriloquist effect occurs when sounds appear closer in space to the visual stimulus than they are in actuality because the locations of the auditory stimuli become biased towards the location of the visual stimuli (Alais & Burr, 2004; Chen & Vroomen, 2013; Warren, Welch, & Mccarthy, 1981). This is thought to occur because spatial resolution is higher for visual stimuli.

In contrast, in the temporal domain, auditory signals were judged as longer than visual signals of the same physical durations (Burr, Banks, & Morrone, 2009). In contrast to the spatial domain, the temporal resolution is greater for audition than it is for vision. As a result, locating the start and end of the auditory signal in time would be easier than locating the start and end of the visual signal. Penney and colleagues (2000) raised the issue that several between-subjects studies failed to show larger estimates for visual temporal stimuli compared to auditory temporal stimuli. They claimed this may occur due to within-subjects designs allowing participants to form a composite memory representation, which included both modalities for reference. Whereas between-subject's participants only have one modality stored in reference memory, depending on which condition they were in (Penney, Gibbon, & Meck, 2000).

To explain this effect the memory mixing model was proposed by Penney and colleagues (2000) and it uses the components of the pace-maker accumulator clock models found in the scalar theory of timing as a basis. These internal clock models will be explored in more detail in section 1.2. The internal clock is able to count more auditory tones than it is visual flickers due to the higher temporal acuity of the auditory system relative to vision (Burr, Banks, & Morrone, 2009; Welch, DutionHurt, & Warren, 1986). This means that auditory tones appear longer than they truly are. Memory mixing is given its name due its claim that reference memory within the pace-maker accumulator models, does not distinguish between modalities. Instead, it states that the memory representations, coming from short term memory and being compared to reference memory, are only defined as being short or long term memory distributions (Gu & Meck, 2011). Auditory tones are more likely to be learned as long durations relative to the standard duration because the number of ticks will be most similar to the 'long' reference memory distributions. Likewise, that visual flickers are more likely to be learned as short durations because they share a more similar number of ticks to the durations found in the 'short' reference memory distributions (Penney et al., 2000)

Magnitude effects

One of the earliest demonstrations of non-temporal magnitude influencing perceived duration was demonstrated by Goldstone and colleagues (1978). When brightness of an LED and loudness of a tone increased, so did the length of the duration estimates (Goldstone et al., 1978). The magnitude of the pitch of the audio stimulus has also been shown to influence the size of participant responses, such that higher pitched auditory stimuli were judged as longer than lower pitched stimuli (Yu, 2010).

Investigations beyond manipulating the audio and visual modalities include the use of vibrotactile stimulation, which led to duration estimates increasing logarithmically with the amplitude of the vibrotactile stimulus (Ekman, Frankenhaeuser, Berglund, & Waszak, 1969). Similarly, when presented with heavier weights, this led to increased estimates of how long the weight was in their hand for (Lu, Hodges, Zhang, & Zhang, 2009).

The impact of magnitude effects could also spread beyond the sensory modalities and act via a general scale of magnitude. The theory positing this scale has been called A Theory of Magnitude (ATOM; Fabbri, Cancellieri, & Natale, 2012) and it incorporates the relationship between number, space and time. It has been argued to account for increases in size (Xuan, Zhang, He, & Chen, 2007) such that larger squares appeared to last longer than smaller ones. Numerosity had a similar effect, whereby arrays of 8 or 9 dots lasted longer than one or two dots (Xuan et al., 2007). It also has the potential to explain the difference in judgements for filled and empty durations. Filled durations are produced using one continuous stimulus whereas empty durations use two clicks or flashes of light to mark out the start and the end of the stimulus (Matthews & Meck, 2016). Filled durations were perceived as longer than empty ones (Plourde, Gamache, & Grondin, 2008). According to the predictions of ATOM, this could occur because we perceive magnitudes as related across time and space. Therefore, if it appears to take up a large space, it will seem to last longer temporally.

Other biases of temporal magnitude estimation

Taken together, the research presented in the ‘Modality effects’ and ‘Magnitude effects’ sections suggests that psychophysical measurement lends itself to understanding the relationship between the physical properties of the stimulus and its perceived temporal extent. More specifically, the research in sections described up until this point has demonstrated how temporal perception may be influenced by the scalar property, the modality of the stimulus and the size of surrounding non-temporal magnitudes. It could be the case that durations were perceived relative to the context in which they were learned, due to the influence of modality and non-temporal magnitude upon the size of estimates. These contextual effects upon duration estimates, imply that our perception of time is not absolute but, varies according to the context in which the time interval was perceived. Historically, magnitude estimation was considered to be influenced by context, for example in the form of regression effects where responses were towards the mean of previously presented stimuli (Cross, 1973; Holland, & Lockhead, 1977; Hollingworth, 1910).

More recently, magnitude estimation processes have shown regression effects and other related contextual effects. For example, during path integration, regression of responses to the mean was shown and the magnitude of the group of stimuli (short, medium or long) influenced the amount of this drift towards the mean (range effect). This was such that larger stimulus magnitudes produced stronger regression (Petzschner, Glasauer, & Stephan, 2015).The final type of systematic biases were found in relation to the sequence of the stimuli. These sequential effects showed the importance of relative stimulus history upon estimates, because estimate sizes increased when the stimuli shown previously were larger relative to when they were smaller (Petzschner et al., 2015).

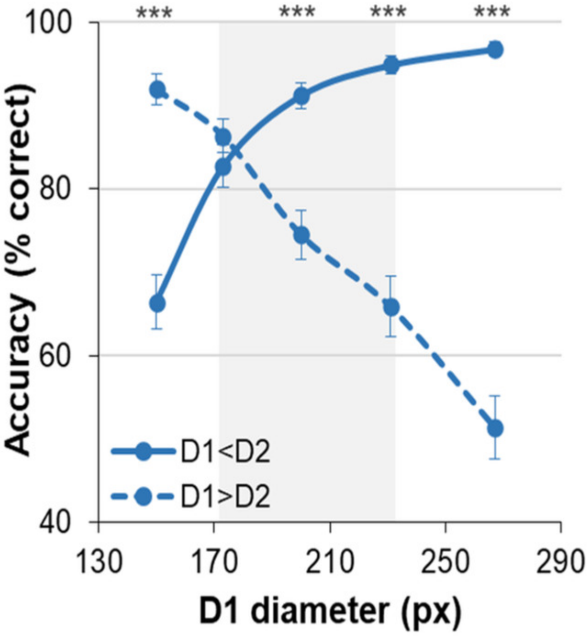
In the area of timing specifically, Vierordt’s law states that when presented with a set of durations, longer durations will be underestimated and shorter durations will be over-estimated, which was the first suggestion of these regression effects within time perception (Lejeune & Wearden, 2009). Regression and other related effects have been studied in relation to timing for example; regression effects (Jazayeri & Shadlen, 2010; Glasauer & Shi 2019), range effects (Hollingworth, 1910) and sequential effects

(Narain, Mamassian, van Beers, Smeets, & Brenner, 2013; Fischer & Whitney, 2014). A distinction that has been made within sequential effects is the influence of the local and the global context upon the estimates, such that the local context incorporated the effect of the most recent trials (where N=1 is one trial previous to the current trial) whilst the global context referred to the range of the stimuli magnitudes used (Hallez, Damsma, Rhodes, van Rijn, & Droit-Volet, 2019). However, it is challenging to study these influences in isolation. For example, in the case of central tendency effects, properties of these effects may be influenced by the global impact of the range of trials used, whereas the formation of these effects may also be related to the local mechanics of trial by trial updating (Glasauer & Shi, 2019).

Biases of magnitude can also be produced when a pair of successive stimuli are presented and compared according to size. One such bias is the time order error (TOE) effect which predicts the over or under estimation of the first duration. Negative TOEs occur when the first stimulus is underestimated relative to the second stimulus and positive TOEs occur when the first stimulus is overestimated relative to the second (Fechner 1860, as cited in Hellström, 1985).

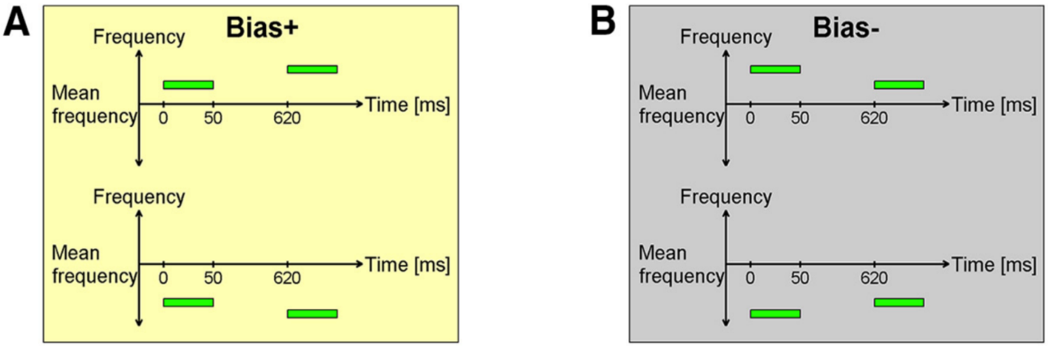
The bias involved in the time order error is often described as a bias towards the centre of the distribution such that the first duration is often overestimated at small values of D1 and underestimated at larger values of D1 leading to worse discrimination for D1<D2 in the former case and improved discrimination for D1<D2 in the latter (see Figure 2). This is because If the bias towards the mean is in a direction that makes the difference between the two stimuli larger, then discrimination performance will be improved. Conversely, if the bias towards the mean is in a direction that reduces the difference between the two stimuli, then discrimination performance will decrease (Jaffe-Dax, Raviv, Jacoby, Loewenstein, & Ahissar, 2015; see Figure 3). The sensation weighting (SW) model is the most commonly used to interpret time order errors and does so by assuming that it is not objective stimulus durations that are compared. It is the comparison of 2 weighted compounds. These weighted compounds incorporate the subjective magnitude of each stimulus and some associated reference level.

Figure 2. Bias in the time order error task



Note. The time order error task involved asking the participants if the second stimulus was larger or smaller than the previous stimulus. Subsequently, the correct responses in a run were summed and then calculated as a percentage of the number of trials in that run. The intersecting pattern on this figure occurred due to the time order error bias. At small stimulus values D1 is overestimated so D1<D2 is initially harder to discriminate, then biased D1<D2 estimates are more likely to be correct as D1 increases in size. For biased estimates where D1>D2, there is a larger percent correct at small values because the over estimation makes this judgement easier, this becomes more difficult as D1 increases. Figure adapted from Sapey-Triomphe, L., Timmermans, L., & Wagemans, J. (2021). Priors bias perceptual decisions in autism, but are less flexibly adjusted to the context. *Autism Research*, *14*(6), 1134-1146.

Figure 3. Biases towards the mean of the stimulus may impact discrimination.



Note. Adapted from “A computational model of implicit memory captures dyslexics' perceptual deficits”, S. Jaffe-dax et al., (2015), *The Journal of neuroscience, 35* (35), pg 12120. Copyright 2015 Society for Neuroscience.

A) The first tone (S1) is easier to discriminate from the second tone (S2) due to the over and underestimation bias. This is because as the perception of the first stimulus drifted towards the mean frequency, it becomes more distinct in magnitude from the second stimulus.

B) The first tone is harder to distinguish from the second tone due to over and underestimation bias. As the perception of the first stimulus drifted towards the mean frequency, its magnitude became closer the second stimulus and therefore discrimination becomes more difficult.

If the bias towards the mean is in a direction that makes the difference between the two stimuli larger, then discrimination performance will be improved. Conversely, if the bias towards the mean is in a direction that reduces the difference between the two stimuli, then discrimination performance will decrease.

1.2 Temporal estimation: theoretical framework and models

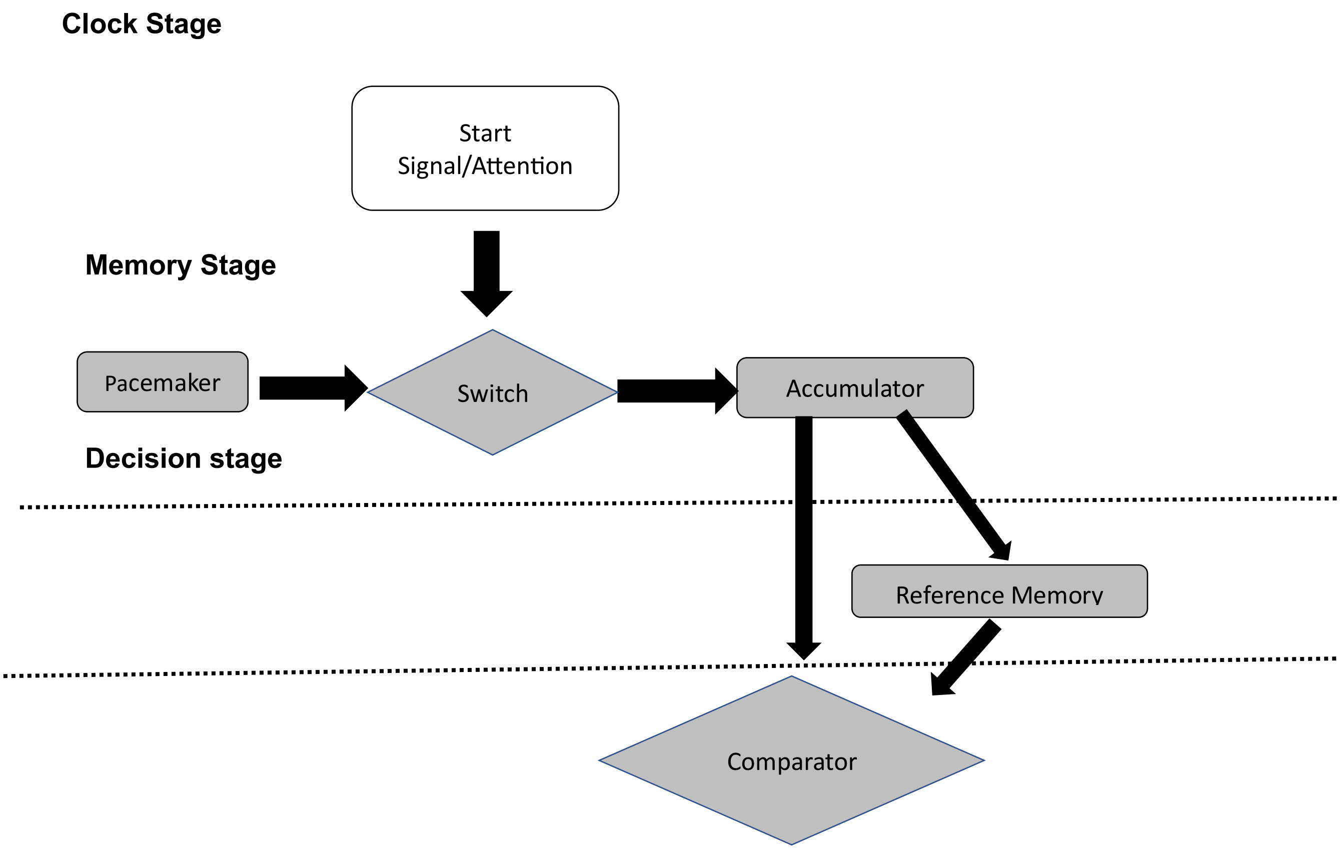
The most well-established model of time perception is the scalar expectancy theory, which proposes that an internal clock mechanism is used to measure elapsed time (Gibbon, Church & Meck1984; Kononowicz, Van Rijn, & Meck, 2016, see Figure 4). The first component of the model is the pacemaker. It produces regular ticks, allowing for a linear measure of time passing because it acts like a metronome (Church, 1984). A switch, or a switch and gate mechanism as suggested by Block and Zakay in 1995, controls when the tics pass through to the accumulator in order to be counted (Zakay & Block, 1995). The accumulator then holds the total pulse count. Finally, when the current interval length is passed through to the comparator it is compared to reference memory (a long-term memory store of all of the previous trial durations). As a result, the model can be broken down into three stages: clock, memory and decision (see Figure 4; Allman, Penney, & Meck, 2016).

Block and Zakay’s (1995) model provided additions to the scalar expectancy theory (SET) model and other models, making it the most recent version of the internal clock model available. The other models which were consulted in the production of Block and Zakay’s (1995) model were Treisman’s (1963) clock model and Thomas’s and Weaver’s (1975) attentional model. Whilst Treisman’s model was an early predecessor of SET, sharing many of the same functional components (Treisman, 1963), Thomas’s model was centred around the role of attention (Thomas & Weaver, 1975).

Other models of time perception have included the striatal beat frequency model, which intended to show how neural mechanisms could account for a timing mechanism in the brain but this representation was not necessarily linear, as it was in the clock models (Grondin, 2010). Striatal spiny neurons are thought to integrate activity from cortical oscillators to produce beats. The beats have periods which span a wider range of durations compared to the original cortical oscillators (Oprisan & Buhusi, 2011). A handful of other models refute the idea of an internal clock or oscillator mechanism entirely and suggested that time is an inherent part of neural firing. Stating that there would be no need to refer to an internal clock because timing is a pre-existing property of the neuro-cognitive processes (Grondin, 2010). One such group of theories includes state dependent networks (SDN’s). These theories suggest that all active regions of the brain are able to encode time in the form of the difference in time between their resting state (t = 0) up until their finishing state of the firing. This difference in time between the two states corresponds to the length of the interval, thus creating a spatial to temporal transformation of the neural firing (Karmarkar & Buonomano, 2007).

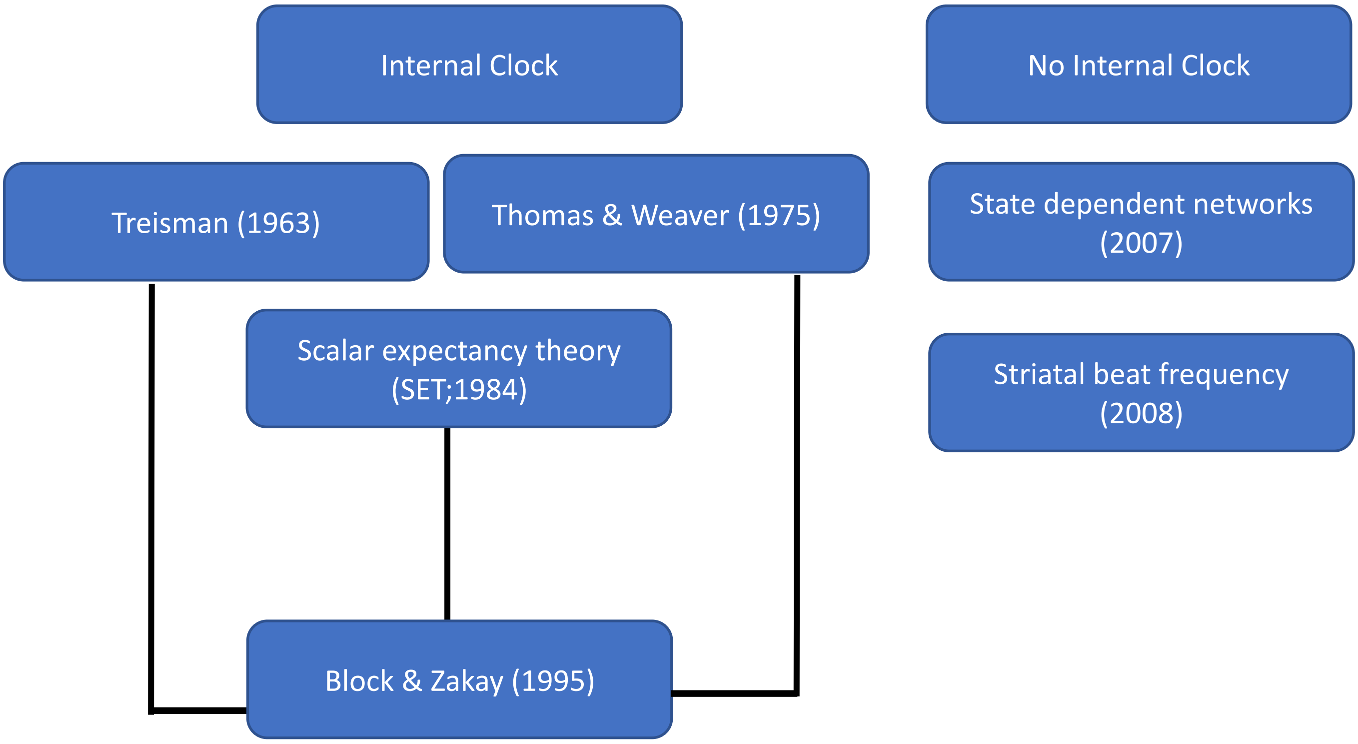
Despite the existence of a large number of models, and their varying success, there is no unified model of time that explains the wide range of findings. The large complexity in the environment means that the brains measurement of time and any limitations accompanying this, will likely take place in multiple brain areas, given the range of potential perceptual and cognitive operations involved in time perception. As described in this section, the internal clock models have provided some theoretical explanations for systematic biases such as changes in clock speed and memory mixing. However, one of the contextual effects mentioned in this section could potentially be explained by the increased weight placed upon prior knowledge within the Bayesian framework of perception. For example, central tendency biases could occur due to the inherent uncertainty contained within sensory receptors and processing when estimating time. As a result, more weight is placed on a Bayesian prior, which in this case, is the mean of the stimulus distribution. The following section will explain how Bayesian inferences can increase reliability of estimates by weighting the relative uncertainty of prior and current sensory inputs. This will lead into how central tendency biases specifically have been explained in terms of a Bayesian framework.

Figure 4. The Pacemaker Accumulator Model.



Note. Diagram is based on the figure displayed in Kononowicz, van Rijn, & Meck, (2018).

Figure 5. Timeline for models of time perception.



Note. Timeline showing the progression of established time perception models chronologically for internal clock and no internal clock models. The connection lines from Block and Zakay (1995) to three other models indicate their contribution to Block and Zakay (1995)'s model.

1.3 Bayesian inference as a theoretical framework for magnitude estimation

The history of the Bayesian observer model

The Bayesian observer model is rooted in Helmholtz's (1876) suggestion of unconscious inference. This unconscious inference as he described, involved automatically perceiving the most likely state of the object of perception, on the basis of experience and images upon the retina (Pellicano & Burr, 2012). Bayesian models build upon this by suggesting that these inferences provide perceptual estimates that reflect the most probable interpretation of the world. The Bayesian observer model has since successfully explained a range of human behaviour, including the previously described contextual effects found in magnitude estimation (Petzschner et al., 2015). It has achieved this by explaining that weights are given to both prior knowledge and sensory input signals when making perceptual judgements. Therefore, it is able to account for experimental findings such as the influence of previously presented information upon current sensory estimates (Glasauer & Shi, 2019).

The need for a mathematical framework

The advantage of this model is that it provides a quantitative explanation of the contribution of prior and current perceptual information. This mathematical framework is necessary in order to quantify and understand how the combination of current sensory signals, and perceptual *a priori* knowledge was possible mechanistically (Rao, Olshausen, & Lewicki, 2002). This also meant that the process of weighting these sensory signals and prior knowledge was now able to be studied, with specific predictions being made about the size of the weights under different conditions (Palmer, Lawson, & Hohwy, 2017). This process provides the value with the smallest uncertainty with the largest weight (Kersten, Mamassian, & Yuille, 2004). Moreover, the quantitative nature of this theory allows for numerical predictions to be simulated. This means that the model has not only explained that these biases exist, but also expanded on this by measuring the extent to which they occur under different conditions of uncertainty.

Where do priors and likelihoods in the model come from?

Consider that you must make an inference about X (the current state of the external world) based on a given sensory input Y. However, it is not the case in human perception that any single input Y maps onto any one specific inference (X) to be detected. Many states of the world could be inferred from the same input. This is due to the level of inherent noise within neurons, and so they may respond differently to the same stimulus. Therefore, all of the different outcomes and their relative probabilities must be considered if an optimal estimate is to be made.

The resulting conditional probability of observing sensory estimate Y given state X of the world P(Y|X) is called the likelihood. When only this probability is used to make predictions, this is the action of the maximum likelihood estimation model. However, there are situations where only taking into account likelihoods, inferred purely on the basis of sensory information, may not produce the most reliable estimate. For example, in environments where sensory inputs are ambiguous, it can be advantageous to also rely on our prior sensory experience. For example, this may involve referring to *a priori* perceptual knowledge such as: I have seen this chair many times, and I know that it was 3-dimensional from the action of sitting. This is where the formation of the prior, or P(X) occurs.

Analysing observer behaviour with Bayesian statistical theory is made possible because Bayes theorem provides a principled way to combine multiple sources of information. According to Bayes theorem, the posterior probability of X given Y is proportional to the product of the likelihood and the prior:

$$P(X|Y) = \frac{P(Y|X).P(X)}{P(Y)}$$

Bayesian inference strategies

As mentioned above the primary goal of the maximum likelihood (MLE) strategy is to maximise the likelihood function that is, the probability of these estimates being produced given the state of the world. It does not incorporate the prior distribution.

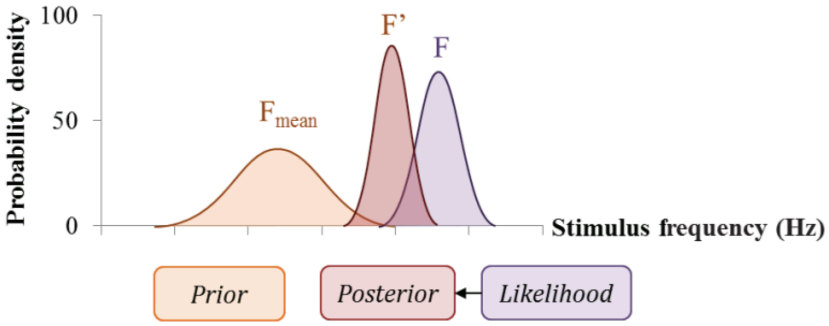
The maximum *a posteriori* (MAP) strategy builds upon the MLE by still acting to maximise the likelihood, but whilst doing so, takes into account the prior. This means that it is able to balance maximising the likelihood function with minimising the cost function, when it is mapping posterior distribution onto estimates. The mapping rule to do this involves using the mode of the posterior and the likelihood. Although, it is only able to penalise all erroneous estimates in terms of how they differ from the most common value and not the average value (Jazayeri & Shadlen, 2010).

A Bayesian inference strategy known as the Bayes Least Squares (BLS) uses a prior and maps the posterior to the estimate using the mean of the posterior distribution. It then specifies a cost function that penalises errors (deviations from the true value) by the square of their magnitude (Jazayeri & Shadlen, 2010).

1.4 Evidence for Bayesian integration in the temporal domain

The following sections will cover the evidence for Bayesian integration in temporal and spatial domains, focusing predominantly on the central tendency effects produced by weighting of the prior. Central tendency effects may be produced where a series of magnitudes estimations become biased towards the mean of the stimulus distribution and therefore displaying the influence of prior knowledge of the stimulus (Jazayeri & Shadlen, 2010; Sapey-Triomphe, 2017). This provided a reason to believe that the Bayesian framework applies here because the current sensory information (likelihood) is influenced by previous statistical regularities in the sensory environment (priors; see Figure 6).

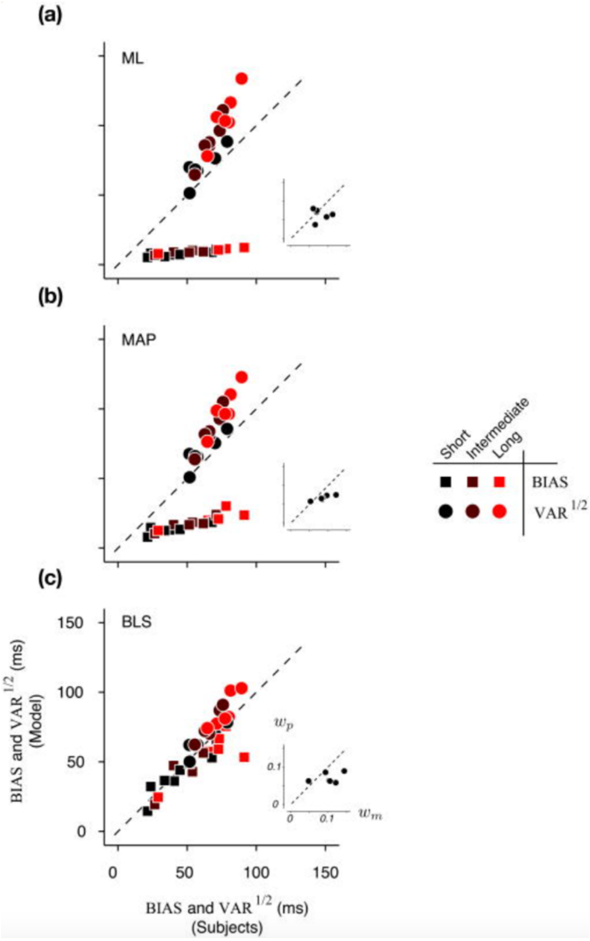
Figure 6. Bayesian priors act to increase reliability when sensory input signals are uncertain.



Note. The likelihood (purple) is combined with the prior (orange). As result, Bayesian integration forms a distribution called the posterior (shown in pink). Overestimation of small stimuli and under estimation of large stimuli indicates a bias towards the mean and therefore a bias towards the prior. where F is the sensory likelihood, F' is the posterior and F_{mean} is the prior. Figure adapted from Perceptual inference and learning in autism: a behavioral and neurophysiological approach by Sapey-Triomphe, L., (2017) , *Doctoral dissertation, Université de Lyon* p62.

Jazayeri & Shadlen (2010) were the first authors to use Bayesian models to examine time perception mechanisms. They investigated how several different Bayesian estimators may reflect participants method of reproducing time intervals. This included the maximum likelihood estimation (where a uniform prior was assumed) and two Bayesian models where the Bayes least squares model and maximum a posteriori model used the mean and the mode of the posterior respectively. Firstly, this research supported the existence of central tendency effects when reproducing time intervals. Then by fitting these models to participants responses, it was found that the Bayes least square model best accounted for the bias towards the mean (Jazayeri & Shadlen, 2010). This meant that using the mean as the prior and the scalar variability to model noise, best explained how participants were making these judgements. Therefore, they not only demonstrated the presence of these central tendency effects, but they also modelled priors and noise in order to understand how these effects occurred (see Figure 7).

Figure 7. The relative success of each model is shown in these plots of the subject's bias and variance$^{1/2}$ as a function of each model's bias and variance $^{1/2}$.



Note. Reprinted from "Temporal Context Calibrates Interval Timing", by initial, M. Jazayeri and M.N. Shadlen 2010, *Nature Neuroscience*, *13* (8), pg 22. Copyright 2011 by PMC.

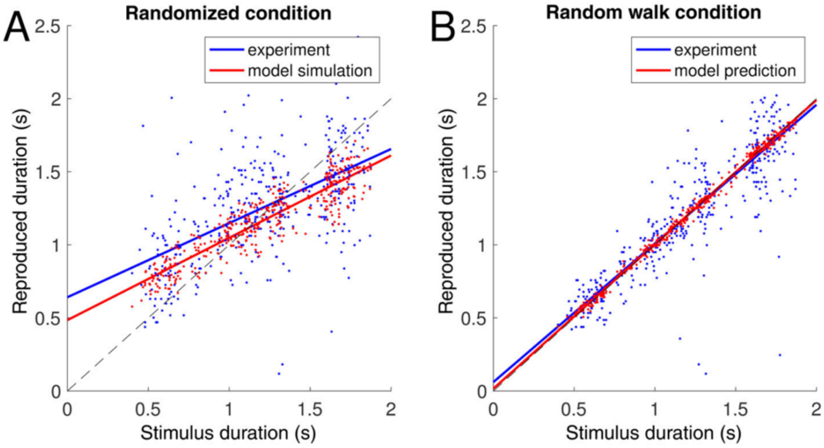
a) the maximum likelihood model in b) the maximum a posteriori model c) the Bayes least squares model. With the bias (squares) and variance (circles) being plotted from 3 prior conditions (short medium and long durations).

All the Bayesian models tested by Jazayeri and Shadlen (2010) incorporated a static prior, which involved setting the prior to be an unchanging value throughout all trials. As a result, the extent to which the prior developed over time was not considered. Therefore, this research allowed for understanding why these central tendency effects occurred that is, to account for uncertainty in sensory estimates, but it was not able to acknowledge how these priors were learned over the course of the experiment. To assess this process of prior updating, an iterative model may be needed because it is designed to understand the influence of trial history on a Bayesian prior. Petzschner (2012) compared static and iterative prior models and found support for an iterative model of distance estimation. It was shown that when compared to a model that used the same prior throughout the trials (the one-stage model as used by Jazayeri & Shadlen 2010), an iterative model (a two-stage model) was better able to explain the data.

The two-stage model was composed of a first stage that was equivalent to the one stage model. Subsequently, the second stage was able to update the prior continually, and therefore considered how the prior changed in order to support the next estimate (Petzschner, 2012). This demonstrated that the iterative quality of the model showed strength in explaining the mechanisms involved in path integration. To gain insight into the influence of the previous trial, the two stage model from Petzschner (2012) uses a Kalman filter. Kalman filters have been used to understand a wide range of estimation behaviour such as, human distance and angle estimation (Petzschner & Glasauer, 2011), reaching behaviour (Verstynen & Sabes, 2011) and numerical estimation (Cicchini, Anobile, & Burr, 2014).

This led authors from the same research group to explore the merits of an iterative Bayesian model in explaining the mechanisms of time perception. Glasauer & Shi (2019) suggested that human perceptual strategies were driven by statistical learning that occurred throughout the duration of the experiment. They used a random walk stimulus sequence (a variation on fully random presentation) to produce small and gradual (one step randomly in the positive or negative direction) magnitude changes in duration. Therefore, the current duration magnitude could be predicted by the previous trial more easily. Random presentation led to large central tendency effects, due to the large amount of uncertainty in the size of the duration that follows the current duration (Glasauer & Shi, 2019). In contrast, for stimuli containing reduced randomness (random walk sequences), a Bayesian Kalman filter model would predict smaller central tendency effects because magnitude changes between each successive trial were more gradual. Their findings supported the idea that ordering stimuli according to a 'random walk' reduced the size of central tendency effects, meaning that priors may be updated via a Kalman filter model (see Figure 8; Glasauer & Shi, 2019).

Figure 8. Central tendency bias reduction for random walk sequences in Glasauer and Shi (2019; 2021b)



Note. Different amounts of central tendency bias were seen in the experimental data (blue dots and linear regression line) and the model (red dots and regression line) for A) randomized stimuli sequences and B) random walk stimuli sequences.

1.5 Evidence for Bayesian integration in the spatial domain

1.5.1 Visual perception of spatial stimuli

Light from above

It has been suggested that a Bayesian prior may be formed in response to how visible light illuminated different surfaces (Adams, Graf, & Ernst, 2004). The process of 'shape from shading' involves inferring the concave or convex appearance (three-dimensional shape) according to how light illuminates the two-dimensional object on the screen (Symons, Cuddy, & Humphrey, 2000).This inference, due to the ambiguous information provided by the incident light direction**,** is influenced by the light from above prior. The stimuli used in Adams and colleagues (2004) may appear as either convex if the assumed direction of incident light is from above and concave if the incident light is assumed to be illuminating the stimulus from below**.** The light from above prior is produced as a result of the sensory knowledge that light sources usually appear above objects. This is often (but not always) the case for natural and artificial light (Adams et al., 2004). Although it is worth noting that natural light from the sun will vary in position according to the time of day.

A Bayesian account would explain that when the sensory likelihood is more uncertain, prior knowledge is used to disambiguate the signals. Prior knowledge would indicate that light sources usually come from above and so this is what is used to form a 'light from above' prior. Although this prior has been sensitive to short term malleability such that haptic feedback allowed for a change in reported direction under ambiguous lighting conditions (Adams 2004). A more recent paper explored how a change in the direction of the incident light and the resulting texture percept (convex versus concave), could be learned according to context (Kerrigan & Adams, 2013). This was such that during test trials red and green spheres were used to indicate a different prior distribution. Participants learned this contingency successfully (red indicated one distribution and green signalled the other distribution) after not having any prior association according to colour during the training trials. Bayesian modelling was able to reveal how learned prior distributions may have a role in responses to different contextual stimulus cues.

Visual orientation.

The oblique effect, present in estimates of visual orientation, is the finding that cardinal directions show more accurate discrimination than oblique directions (Appelle, 1972).This effect is thought to occur because cardinal lines present themselves more commonly in nature (Coppola, Purves, McCoy, & Purves, 1998)**.** Bayesian inference could be a potential mechanism for this because cardinal orientated lines are more frequently perceived than oblique directions in natural environmental statistics. The

Bayesian observer model would explain that a stronger prior would be associated with sensory aspects that occur more frequently in the learning environment (Adams et al., 2004). Research examining this possibility showed that participants displayed biases towards the cardinal directions and, this was evident in their prior distributions (Girshick, Landy, & Simoncelli, 2011). Overall, their research implied that a Bayesian model was able to account for a bias towards the statistics of natural environments. Although the necessity of priors in explaining this difference can be debated due to the selectivity of V1 orientation neurons for cardinal directions (Furmanski & Engel, 2000).

Visual motion

Biases in other fundamental aspects of spatial perception, such as visual motion perception, have also been explained by the process of Bayesian integration**.** For example, Weiss, Simoncelli and Adelson (2002) concluded that biases in speed perception occur as result of optimal Bayesian perception. This likely occurred because human measurement of motion is noisy (due to inherent sensory noise), so prior knowledge was weighted more heavily. This prior reflects the statistics of the natural environment where slow speeds are more frequent that high speeds, therefore biases towards slower speeds occur as a result of a Bayes-optimal judgement (Weiss, Simoncelli, & Adelson, 2002). In another experiment, Stocker and Simoncelli (2006) presented moving gratings to participants, and they had to respond according to which of the 2 gratings was faster. Whilst other Bayesian models existed at this time, their model was unique in that it accounted for the internal noise of measurements. This meant that it accounted for trial-to-trial variability in the likelihood distribution. Overall, they showed a good fit to the data and also performed well compared to other models, which had many more free parameters (Stocker & Simoncelli, 2006).

1.5.2 Self-motion cues in the spatial domain

Path integration

The Glasauer group has used a virtual reality environment to explore Bayesian prior updating during path integration. Participants were moved through a virtual environment either by propelling themselves a certain distance through the environment or via rotation in the same spot. The existence of Bayesian integration was supported by a regression towards the mean of the distribution of distances, and of the orientations they had experienced during the experiment (regression effect; Petzschner & Glasauer, 2011). Moreover, the 'large' set of stimulus magnitudes gave rise to larger biases than the shorter set of stimulus magnitudes (range effect; Petzschner et al., 2015). Finally, as the mean duration reproduction length increased, the standard deviation of participants reproductions increased (Petzschner & Glasauer, 2011).

Computational modelling was used to fit a Bayesian model (where the mean and standard deviation was used to form the prior) and a simple iterative model (two stage model) allowed for the prior updating process to incorporate the prior on the previous trial into the current prior. The model fits suggested that the central tendency effects seen in the empirical data were best described by a simple iterative model, as opposed to a static model. Therefore, the origin of the central tendency bias in this case was shown to have a prior updating mechanism that uses information from the most recent trial. More details on how these models, and models from other research groups, will be expanded on in a later section of the literature review. Overall, this research demonstrated that central tendency effects in the path integration can be explained by a simple iterative Bayesian model.

1.5.3 Cue combination (visual and auditory stimuli in the spatial domain)

Combining cues from multiple senses is incredibly important for navigating a 3D environment. Studies have shown integration of visual and haptic cues (Ernst & Banks, 2002), texture and motion cues (Jacobs, 1999), audio and visual cues with both simple stimuli (Alais & Burr, 2004; Körding et al., 2007) and speech stimuli (Arnold, Tear, Schindel, & Roseboom, 2010; Bejjanki, Clayards, Knill, & Aslin, 2011). For integration of visual and audio cues this often involves presenting the audio and the visual signals in varying locations and asking participants where they perceived the audio or visual stimulus to be located. Then it can be said that audio visual integration has occurred when the audio and visual stimuli are perceived as possessing the same source location.

One such experiment by Körding and colleagues (2007) aimed to determine which model could best explain this cue combination behaviour. The experiment involved displaying simple visual stimuli on the screen and auditory stimuli through headphones. Using this equipment, they were able to vary the spatial location of the stimuli either in terms of their position on the screen and loudness through the headphones. Participants were able to respond using two rows of buttons to indicate the perceived location of the audio and visual signals (Körding et al., 2007).

They proposed a Bayesian causal inference model, and it was able to predict 1) participants behaviour during cue combination more successfully than previous models and 2) whether single or multiple sources of the sensory signals would be perceived. The Bayesian element of this model indicates how much we can trust each signal according to its relative uncertainty and provides information about about joint or separate causes. When perceiving speech stimuli, a Bayesian explanation has been supported because they made decisions according to sensory uncertainty that were in line with a Bayes optimal observer (Bejjanki et al., 2011). Therefore, there is a clear advantage in being able to combine visual cues with cues from other modalities (sound), according to their uncertainty, when trying to reliably determine the location of something that could be the source of both cues.

Up to this point, the current section has covered the combination of two sensory cues. However, it may also be possible, during the estimation of sensory magnitudes, for sensory information to be combined with information from symbolic verbal cues (Petzschner, 2012)**.** Three different behavioural conditions were used to determine whether the integration of sensory and verbal cues was possible. The first two out of the three conditions acted as controls because their purpose was to intentionally produce scenarios where: 1) the stimuli appeared as two separate blocks, one with short intervals and one with long intervals or 2) the long and short intervals were presented in the same block.

The third condition was similar in presentation to the second condition (interleaved instead of blocked), but it also involved the participants being verbally cued before each stimulus presentation as to whether it was going to be a short or a long stimulus.

If the participants data appeared like the first control condition, then this finding would have supported the idea that the symbolic verbal cue was integrated into their estimate. However, if the participants still treated the intervals as one continuous range, as they demonstrated in the second condition, then the symbolic verbal cue was not integrated. This research supported the integration of the symbolic cue with the sensory cues because participants were able to distinguish long and short distributions without the need for separate blocks.

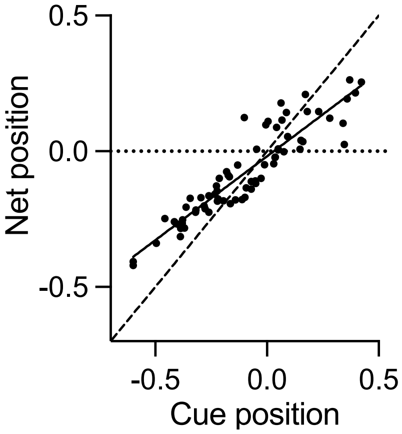
There have also been demonstrations of the combination of two cues from the same sensory modality to improve the confidence of sensory judgements (Bejjanki, Knill, & Aslin, 2016; Berniker, Voss, & Kording, 2010; Kiryakova, Aston, Beierholm, & Nardini, 2020). The experimental paradigm of spatial localisation, which will be elaborated on in the following section, aimed to understand how visual cues may be combined with previous knowledge of the location of a target stimulus (prior). An experimentally manipulated amount of noise was necessary in the spatial location of the visual cue due to the ability of the human eye to estimate spatial positions with high accuracy. During spatial localisation, the visual cue (noisy indicator of the target) and prior (where the target location was revealed to be on previous trials), may be combined according to their relative uncertainty using Bayesian inference (Berniker et al., 2010).

1.5.4 Evidence for Bayesian integration in spatial localisation

Spatial localisation research has presented two virtual coins, one after another to see whether humans use Bayesian inference to estimate their location. This is known as the ‘Catch The Coin’ task (Tassinari, Hudson, & Landy, 2006). The task involves a cue coin being presented first, shortly followed by a net which needs to be dragged by the participant to where they think the hidden target may be. Finally, the position of the hidden coin is revealed, and it provides to some extent, prior knowledge of the hidden coin position for future estimates.

The cue coin positions, however, formed the likelihood because they provided current sensory knowledge at the time of responding. This is because the cue coin positions were drawn from a Gaussian distribution (a normal distribution where the mean was equal to the target coin on that trial) and so acted as a noisy cue for the actual target position. According to Bayesian theories, when the standard deviation (variability) of all the possible cue coin positions is large, the weight given to the cue coin will reduce and the weight given to the prior will increase (Bejjanki et al., 2016; Berniker et al., 2010).

Figure 9. Linear regression for spatial localisation



Note. The black data points represented the cue coin position (x co-ordinate on the screen and y=0) on each trial plotted as a function of where participants placed their net response (x co-ordinate only).
The line is a fitted linear regression of this function, where the gradient of the slope was equal to the amount of bias towards the centre of the target distribution.

Initial research using the catch the coin task (Berniker et al., 2010) demonstrated that participants weighted the prior and likelihood according to their uncertainty. More specifically, as the prior variance decreased (became more certain), the participants net position moved towards the prior mean. Whereas when the likelihood variance was lower, participants beliefs about the target location drifted closer to the cue coin (likelihood), meaning that they were weighting the likelihood more heavily. This was shown visually using a linear regression where a decreased slope in the line indicated a move towards weighting the prior and a increased (steeper) slope conveyed a move towards weighting the likelihood (see Figure 9). Overall, Bayesian integration was likely to have occurred because the likelihood distribution (cue locations) was combined with the prior (prior knowledge of the target location) to estimate the current location of the target.

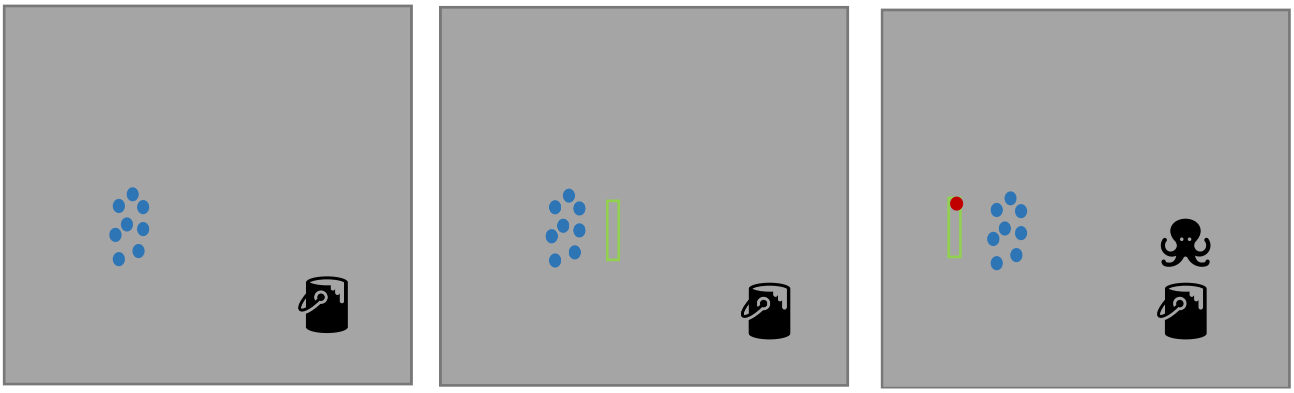
In experiment 1 they manipulated prior variance (participants were presented with a prior variance of 0.05 or 0.2) whereas in the experiment 2 the prior variance switched between two values (one prior variance was chosen randomly at start and after half of the trials had been presented, the prior variance switched). For experiment 3, the target position distribution could be centred on one of two positions (after 10 trials the mean would change, and this pattern repeated for 50 trials). To account for this, they were told at random times the imaginary person throwing the coins will move.

The findings from experiment 1 and 2 demonstrated that irrespective of whether prior variance was constant, or it switched half-way through the blocks (from 0.2 to 0.05), participants learned the prior mean very quickly. They also learned the prior variance, but over a slower timescale closer to 200 trials (Berniker et al., 2010). Experiment 3 aimed to understand if participants could learn as prior means changed independently of variance (variance remained constant). As in experiments 1 and 2 participants learned both prior means quickly whereas they found it much more difficult to learn the prior variance than in the previous experiments. This could have been due to the frequent switching of the prior mean, combined with the variance for both prior means being within one SD of the likelihood and the prior (Berniker et al., 2010). Taken together this means that the prior mean can be learned in various contexts, even as the mean or variance of the target stimuli distribution changes.

Berniker and colleagues (2010) simulated the experiment thousands of times with a human net response being replaced with the model response. It was assumed that the model could accurately know the likelihood variance. The Bayesian model was then compared to participant responses from the first experiment in order to examine the difference between them. There was a good fit to the data in experiment 1 and subsequently, the model successfully predicted the results of experiment 2. Whereas a simple non-Bayesian model, which used a constant learning rate, demonstrated a much poorer fit to the data in experiment 1. Then for experiment 3 this model was adapted to estimate 1 of 2 possible prior means. The model knew both the location of the two means and the correct possibility of switching between them when making estimates. In contrast to the previous experiments, participants found it more challenging to make estimates of the prior variance in experiment 3. It is likely that regular switching interfered with participant estimates because the size of the prior variance (uncertainty) was considered the source of the variable prior mean (and not the existence of 2 distinct means), leading them to over-estimate the certainty of the likelihood. This increased weighting of the likelihood, relative to optimal Bayesian weighting was shown in the data for experiment 3.

However, this earlier exploration of spatial localisation used only single Gaussian distributions. Complex generative models based on several Gaussian distributions would be more likely to reflect natural environments (Bejjanki et al., 2016). Bejjanki and colleagues (2016) set two different prior distributions (bimodal) to form the complex generative model. The target position on any one trial would be drawn from the two Gaussians, always centred on two different locations. To create this in the stimuli, the two target coin positions were produced from both x and y co-ordinates and the sensory likelihood was shown in the form of a dot cloud (see Figure 6). The dot cloud had a set variance that produced variability (low, medium or high) in the distribution (likelihood variance). The dots in the cloud were drawn from a distribution that was centred on the true target location. It was shown that participants were able to learn the prior means of the two different Gaussian distributions (always centred on 2 different locations of the screen) and the prior variances but the variances were learned more slowly than has been observed with single prior distribution models. Therefore, Bayesian integration can occur in more complex environments where there are multiple distributions to be learned.

Figure 10. Stimulus presentation for Kiryakova et al., (2020).



Note. The sequence of stimuli presented to a participant on a single trial in Kiryakova et al., 2020 starting with a dot cloud 'splash' (likelihood), followed by a rectangular net controlled by participants and culminating in the position of the target being revealed. The target was represented by a red dot and if the participants were correct an animation of an octopus falling into the bucket would be shown. This is a representation of the stimuli used in Kiryakova et al.,2020.

Kiryakova and colleagues (2020) explored variability in the likelihood by using a dot cloud (representing the sensory likelihood) and referred to it as a splash in the lake.

The procedure operated similarly to Berniker and colleagues (2010) in relation to the ‘catch the coin’ goal of the task but with a few modifications (see Figure 10). Their aim was to investigate transfer of Bayesian weights to new contexts. Transfer in this context refers to how participants may be able to immediately (without any training at this new likelihood variance) give less weight to a more uncertain likelihood. The authors were motivated to use Bayesian transfer due to its ability to distinguish the existence of Bayesian mechanisms from a reinforcement style learning of prior and likelihood pairs.

In experiment 1 the authors tested transfer to higher variability likelihoods. To achieve this, participants were trained on likelihoods with low and medium variances and then subsequently experienced testing blocks with high variances. Bayesian transfer predicts that they should have weighted this new likelihood to a lesser extent than the previously learned likelihood (as a result of it being more variable than training likelihoods even though they had not fully experienced it yet). There was weak evidence for this effect because the participants were suboptimal in their weighting of the new likelihoods (high variance) by weighting them more heavily than would be predicted by Bayesian models.

The second experiment was identical in procedure, other than the addition of an explicit instruction of different variabilities in the prior. This was done by saying “one of the octopuses moves around a lot and the other stays still”. It was found that there was some evidence for Bayesian transfer because participants did place less weight on the new and more variable likelihood that is, from post-hoc analyses there was a significant difference in the weight placed on the likelihood when the trials incorporated low likelihood variance compared to the medium and high likelihood trials. They also produced closer to optimum likelihood weights as shown by the graphs of likelihood weight but did still not reach the recommended optimum weights. Experiment 3 was a purposeful replication of Bejjanki and colleagues 2016, this meant that all likelihood variances (low, medium and high) were presented throughout all blocks. Transfer was not relevant here as they wanted to see how optimal participants could be in the absence of transfer. They continued to make less optimal estimates than would be predicted by Bayesian models (Kiryakova et al., 2020). The participant estimates remained suboptimal because they placed too much weight on the likelihood. This suboptimality has been demonstrated multiple times in research (Bejjanki et al., 2016; Kiryakova et al., 2020) and the reasons for this may include participants taking into account internal noise of their own estimates (Kiryakova et al., 2020).

1.6 Models of magnitude estimation: static

Generative models provide the basis for different prior updating strategies because they incorporate assumptions about how stimuli are generated, and how prior acquisition should respond as a result. Current research has described static models as assuming the stimuli were drawn randomly from the distribution, with a mean and the standard deviation that are relatively stable. Whereas iterative models assume that the stimulus fluctuates through small random variations over the time course of the experiment. Most recently the two state model (Glasauer & Shi, 2021a) has begun to include the assumptions of both models. A review of these models and their key findings will be detailed in this section.

Static Bayesian models typically assume that the prior is learned from the distribution of the stimuli (Glasauer & Shi, 2021b). One of the early static models was used in a temporal reproduction experiment conducted by Jazayeri and Shadlen (2010). In this case, the priors had a uniform distribution meaning that the prior was assumed to be a one to one mapping of the entire stimulus distribution and not a Gaussian shaped approximation of the stimulus distribution based on the mean and standard deviation. In Jazayeri and Shadlen (2010), it was assumed that participants learned the distribution, and this was reasonable, given they trained the participants on the distributions using thousands of trials. However, the question of how priors were learned over time remained and this will be covered in the following section.

Jazayeri and Shadlen (2010) found a clear central tendency bias (over and underestimation) was shown, and this was such that the central tendency bias was stronger for longer durations compared to short durations. The authors then fitted several models including a Maximum Likelihood (MLE; which uses sensory likelihood and no prior, therefore the most likely value). Then they also fitted the maximum a posteriori (MAP; this model uses the peak of the likelihood) and the Bayes Least Squares (BLS; which mapped the mean of posterior distribution onto the estimate) models. The results demonstrated MLE to be a poor fit because it didn't take prior dependent biases (central tendency) into account. The MAP provided a slightly better fit of the variance and bias. Finally, the BLS model demonstrated the best fit of the bias and variance. The improved fit of the Bayes least squares model could also not be attributed to over fitting because all 3 models each had 2 free parameters.

Cicchini and colleaugues (2012) also fitted Bayesian models to temporal estimation data by assuming a static prior. They measured temporal reproduction and temporal bisection (a task which involves training participants to learn the pairing of a certain stimulus with a short duration and one with a long duration. Then on subsequently un-trained durations they must select which stimulus they thought was longer or shorter in line with the trained pairing. The number of 'long' responses were recorded) on non-musicians and two types of musicians (drummers and bow string players). In the case of auditory stimuli, responses were highly accurate in all groups, providing support for the increased temporal resolution of the auditory system and as a result increased accuracy.

For the visual stimuli, the non-musicians and the string musicians showed a clear central tendency bias, similar to the levels seen in Jazayeri and Shadlen (2010). However, the drummers demonstrated almost veridical temporal reproductions. By partitioning the error in the reproductions from these groups, the authors were able to show increased bias in the line with a measure of individual variability (coefficient of variation). This indicated that the central tendency bias provided a worthwhile compromise in accuracy when variability of responses was high (non-musicians and string musicians). However, when variability was low, in the case of the drummers, central tendency bias was reduced because any advantage provided by the bias was no longer necessary under less variable (more certain) conditions. This occurred because there was less reliance on the prior and sensory information was more precise (Cicchini, Arrighi, Cecchetti, Giusti, & Burr, 2012). They fitted Jazayeri and Shadlen (2010)'s model which assumed that the prior comprised of a uniform representation of the entire stimulus distribution. In addition to fitting a uniform distribution they also fitted their own model that used the standard deviation and the mean to represent learned characteristics of the distribution as parameters for the Gaussian curve prior. The uniform prior model used in Jazayeri and Shadlen (2010) showed some ability in predicting increased bias for increasing variability, however their own model (Gaussian curve) provided a better model fit.

Roach and colleagues (2017) continued to assume a static prior distribution. Although a Gaussian curve was selected over a uniform distribution as it was suggested to provide a better fit by Cicchini and colleagues (2012). The authors modelled temporal reproduction estimates to explore prior acquisition in different contexts. For instance, they presented subjects with interleaved presentations of long and short durations and a single prior distribution was learned. In contrast, blocked presentation of each duration type (short and long) produced distinctly different priors. For different motor response types, including button press and speech based responding, two different prior distributions were shown, even when the two response types were interleaved in a single block. This meant that information about the stimuli could be incorporated into the same prior distribution, but this was not true when they were paired with a different motor output (passive viewing, active reproduction or reproduction of a speech sound such as ‘ba’). Taken together, priors may be formed generally when there are produced from sensory input and formed more specifically when they are learned for different motor output.

As shown in earlier models, the authors assumed a static Bayesian prior model when fitting participants responses. A logarithmic relationship between the perception of time and physical time was also assumed and as a result, the model produced the inherent scalar property of timing. For conditions where participants were presented with interleaved short and long durations the model fits demonstrated that participants assumed a Gaussian shaped distribution where the mean and standard deviation were estimated. More specifically, one large prior was learned rather than short and long as would be seen in blocked presentation. This was emphasised by the fact the mean of their prior existed in the medium range, where no stimuli were presented. This meant that participants may initially learn an approximation of the entire distribution (a prior mean and standard deviation), before being trained on all possible magnitudes within that range (Roach et al., 2017).

1.7 Models of magnitude estimation: iterative

In the case of iterative models, the properties of the prior update according to the most recent stimuli. In doing so these models consider how priors must update over time to successfully provide information about short term changes to the stimulus. This means that the model has different assumptions about how the stimuli were generated (generative assumption) and as a result, a different prior updating process in response to the assumptions it makes about the properties of the stimuli.

Petzchner and Glasauer (2011) found that central tendency effects were produced in angle and distance estimation in a virtual environment. They also found that the central tendency effect size increased with the size of the stimulus magnitude. A further finding was that as the mean duration reproduction length increased, the standard deviations of the reproduction length also increased such that longer distances showed more variable responses. In addition to the empirical data collected during angle and distance estimation, the authors were also comparing the fits of a static model (where the mean and standard deviation was learned) and a simple iterative model (a Kalman filter). The Kalman filter was able to iteratively update the prior by incorporating 1) the prior on the previous trial and 2) a certain level of random noise associated with trial to trial variation.

The iterative and static models were set up using Bayes rule and the assumption that the internal noise of magnitude estimates was logarithmic. Therefore, in a given trial, the stimulus duration (d_m) was assumed to produce a log spaced internal representation and that the internal representation of the stimuli (x_m) would be equal to the natural log of the duration divided by a constant (d_o) plus some additional noise (n_m).

$$x_m = \ln\left(d_m/d_o\right) + n_m$$

The noisy sensory inputs can be integrated with a Bayesian prior (prior mean, prior SD) to improve estimates. Bayesian integration then determines that the sensory likelihood and the prior would then produce the posterior distribution with the following equation:

$$x_r = w_{prior} \, . \, x_{prior} + w_m \, . x_m$$

The weights used as part of Bayesian integration account for the uncertainty associated with the likelihood and prior because they use the inverse of the variance (precision):

$$w_{prior} = \frac{1/\sigma^2_{prior}}{\frac{1}{\sigma^2_m} + \frac{1}{\sigma^2}_{\ prior}}, w_m = 1 - w_{prior}$$

Iterative model updating

The iterative model updates its prior over time via the Kalman filter process and so the mean of the prior at a given trial (i) is:

$$x_{prior,i} = (1 - k_i) \, . \, x_{prior,i-1} + k_i \, . x_m$$

Where the Kalman gain is:

$$k_i =. \frac{\sigma^2_{prior\ j-1} + \sigma^2_{sys}}{\sigma^2_{prior\ j-1} + \sigma^2_{sys} + \sigma^2_m}$$

Additionally, where:

σ^2_{sys} is the variance of the system noise.

$\sigma^2_{prior,j} = k_i \, . \, \sigma^2_m$ is the variance of the prior.

Petzchner and Glasauer (2011) used computational modelling to show that central tendency effects were best explained by the iterative updating of the prior in line with the current stimulus. They found that an iterative model provided a better fit than a static model. The iterative model was more able to account for small variations in participants' data over the course of the experiment. This could be because it included the sensory measurement of the most recently presented stimulus within its prior, whereas a static model did not (Petzschner & Glasauer, 2011). This means that relative to models that initially assumed a static prior distribution, a model that iteratively updates the prior distribution over time was more successful in describing human behaviour in distance and angle estimation.

Subsequent studies conducted by Glasauer and Shi (2019;2021b) aimed to re-evaluate how task order relates to central tendency effects. In this case, they were referring to the central tendency effects described by Vierordt's law as they studied temporal reproduction specifically. It is possible that the random sequences used in stimulus presentation may have been responsible for central tendency effects. In contrast to experimental conditions, It is likely that many changes in the real world occur gradually between a relatively constrained range of values. For example, the change in speed of a moving car is gradual and continuous and not produced by sporadic fluctuations in speed. In light of this, they predicted that an iterative model like those described so far may be more accurate in describing human magnitude estimation behaviours than a model that learns only the characteristics of the distribution (static). To explore these differences, they conducted a temporal reproduction experiment with two different types of stimulus sequence: fully random and random walk.

Random walk stimuli, produced by a Wiener process in the case of this experiment, follow an autocorrelated sequence that may better reflect the gradual changes that are seen in nature. The stimulus sequence produced by this process contained temporal duration magnitudes with a small amount of variation from trial to trial. During random walk sequences an iterative model predicts reduced central tendency effects**.** This is because an iterative model is well suited to estimating the small amount of trial to trial variability present in the random walk sequence specifically. Whereas for a static model, there would be the same amount of central tendency for fully random and random walk conditions because this model does not predict anything in relation to the stimulus sequence.

To investigate their hypothesis that central tendency effect arose due to random stimulus presentation, they simulated the original data (Glasauer & Shi, 2021b). This simulation demonstrated the ability of their model to reproduce data almost identical to the participant data collected by Vierordt in 1868. Subsequently, they modelled the data to show what responses to random walk stimuli would have looked like. As predicted, there was a very large decrease in central tendency bias in the iterative model for random walk sequences. This corroborated their prediction about the importance of experimental protocol.

Glasauer and Shi (2019;2021b) also fitted the static and iterative models described in (Petzschner & Glasauer 2011) to their empirical data. For this experiment there were two highly relevant assumptions of iterative two stage model in relation to the amount of central tendency expected. Firstly, some serial dependency that is, the error in the current trial (k) depends on the stimulus difference between trial k and k-1. Secondly that the strength of the central tendency effect depends on the order of the stimuli (Glasauer & Shi, 2019, 2021b). The iterative model assumes a certain amount of error depending on the trial to trial variability (order) of the stimulus. If stimulus differences are small and serially dependent, the iterative model predicts the size of the error will be small and the amount of central tendency bias will be reduced in random walk conditions (see Figure 8).

Taken together, the empirical findings demonstrated a reduction in central tendency for the random walk condition indicating that the iterative model could be contributing to this reduction in uncertainty. Additionally, the iterative model provided the best fit for both types of stimulus condition Glasauer and Shi (2019; 2021b). This was surprising considering a static model may have been better suited to a fully random stimulus sequence. The success of the iterative model emphasised that by investigating how priors change over time, unexpected mechanisms of human perception could be revealed.

1.8 The Two-State model

Generative models contain assumptions about how the stimuli were generated. The construction of the two state model involved combining the generative models of static and iterative prior updating processes. This is because like a static model, it holds the assumption that any given stimulus (x) was drawn from a distribution with a certain mean ($\bar{x}$) and standard deviation that can be approximated based on a series of stimulus observations:

$$x = \bar{x} + \varepsilon_x$$

Where ε_x is the amount of noise associated with the process of drawing a single stimulus from the Gaussian distribution.

In line with the assumptions of iterative models, the two state model also assumes that there is some small random shift to the values of the stimulus from trial to trial (see Figure 11). More specifically, in purely iterative models, the current stimulus (x) is thought to be equivalent to the previous stimulus (x_{i-1}), with some random change (ε_m):

$$x = x_{i-1} + \varepsilon_m$$

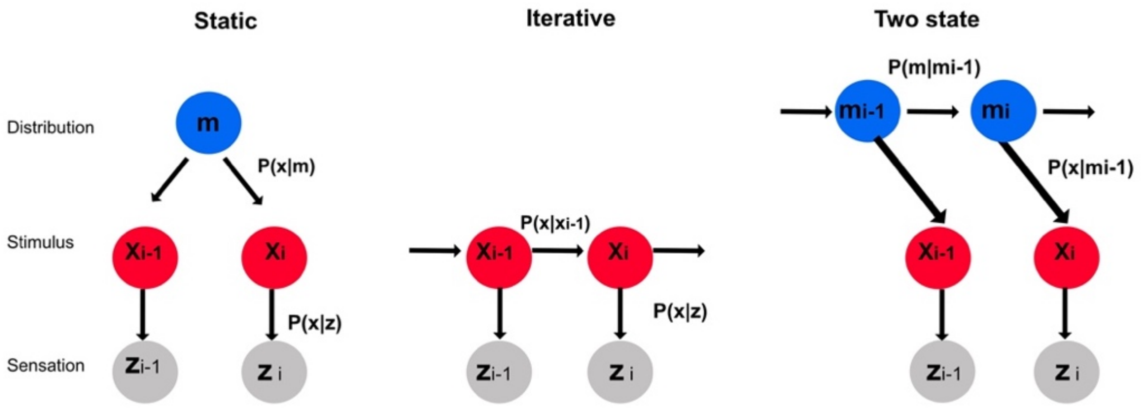
The two state model therefore uses generative assumptions from both static and iterative models to combine their abilities (see Figure 11). As a result, the two-state model has two equations governing the generative assumptions:

$$\bar{x} = \bar{x}_{i-1} + \varepsilon_m$$

$$x = \bar{x}_{i-1} + \varepsilon_x$$

The outcome of this is that the generative model now has two states 1) the randomly changing mean of the stimulus distribution as in the first equation 2) the current stimulus that is assumed to be drawn from this distribution, as in the second equation (see Figure 11).

Figure 11. The assumed generative models and how they access sensations from the stimulus.



Note. The static model assumes that any given stimulus is drawn from the distribution P(x|m) with a fixed mean and standard deviation.

The iterative model assumes that the current stimulus is equal to the previous stimulus P(x|x $_{i-1}$) with some random noise associated with the shift from the previous stimulus to the current stimulus value. There are no assumptions about the mean of the distribution.

The two-state model assumes that the mean of the stimulus (m) may shift and that this shift in the mean from trial to trial then produces a second state whereby the value of the current stimulus is drawn from a shifting distribution. The current stimulus is therefore equal to the previous stimulus with some noise determined by the amount the mean has shifted.

In research conducted by (Glasauer & Shi, 2019) iterative models provided a better fit to the data than static models because they predicted a reduction in compression magnitude (central tendency bias) in response to random walk sequences. This reduction was shown in participants data, but it was a smaller reduction than that predicted by the iterative model. Initially, the reason for this was unclear. Glasauer and Shi (2021a) used a two state model on the same temporal reproduction data and demonstrated a better approximation of the participants reduction in compression than the iterative model. Therefore, a two-state model provided a better fit for participant data relative to static or iterative models. In the current research, a two-state model will be fitted in addition to static and iterative models with the aim of understanding the merits of this type of model. Further detail from the Glasauer and Shi (2021a) will also be provided in Chapter 2.

1.9. Development of magnitude estimation abilities

1.9.1 Time perception and development

The neurotypical development of time perception

Early research into the development of time perception has shown that the variability of estimates decreased with age (Goldstone, Boardman, & Lhamon, William, 1958; Goldstone & Smythe, Elizabeth, 1957). According to research using a child friendly temporal bisection task, the sensitivity to time intervals as measured by the steepness of psychophysical functions also increased with age. The steeper functions indicated a decrease in variability and hence an increase in sensitivity (Droit-Volet, 2003). Upon further inspection, there appeared to be stages of this development such that 5-year olds showed lower sensitivity to intervals than 8-year olds (Droit-Volet, 2003; Droit-Volet, Clément, & Fayol, 2003; Zélanti & Droit-Volet, 2011) and that children overall showed less sensitivity than adults (Droit-Volet, Meck, & Penney, 2007). Although it has been argued that 9-year olds may show equivalent sensitivity to adults at short durations (Zélanti & Droit-Volet, 2011). Furthermore, it has been found that between the ages of seven and eight may be a marked transition phase, where the level of variability in children’s estimates becomes more adult like (Espinosa-Fernández, de la Torre Vacas, García-Viedma, García-Gutiérrez, & Torres Colmenero, 2004).

It has been proposed that attentional capacity, and how it develops, is likely to be the reason for the developmental patterns described so far (Droit-Volet & Gautier, 2002). For intervals up to 2.5 seconds, 9 year olds showed adult levels of performance, demonstrating how children’s time perception may be similar to adults once the attentional processes for longer intervals have been refined (Zélanti & Droit-Volet, 2011). Therefore, certain attentional factors, such as the length of time they were required to attend for, may be able to modulate this low sensitivity in children. Research has shown an attentional modulation effect when a 'click' cue preceding the stimulus was used to capture attention. It was found that sensitivity improved for the youngest groups (3 and 5 years old) as they received the most benefit from the 'click' cue**.** Taken together, this research emphasised the need to account for children's attentional capacity, as determined by their age, when researching the differences in temporal sensitivity.

The tasks used to show the decrease in children's variability as they aged usually involved temporal bisection and or temporal reproduction. The temporal bisection task was used to measure the discrimination of, and hence sensitivity to, time intervals. As part of a pre-training phase participants were asked to view a short and long interval (Rattat & Droit-Volet, 2001). For the benefit of the children, the experimenter would indicate the size of the interval being presented by saying "look it's the short circle, it stays on for a short time" and adjust this speech accordingly for the long interval. They were then presented with a series of intervals that included the long and short trained intervals, along with several intermediate durations. Participants were then asked to report whether the duration they were presented with was either closer in size to the long or short duration. The training phase involved feedback as to whether they were correct (with the use of a clown image when the participants were children) and the final test phase was conducted in the same way but without feedback (Droit-Volet & Wearden, 2001). A psychophysical function was then produced by plotting the proportion of times the 'long' response was made against the true length of the stimuli durations. The bisection point was the point where the x axis that corresponds to half-way along the y axis. It indicated the duration at which it was equally likely that participants were to make the 'long' response as they were to make the 'short' response.

On the other hand, temporal reproduction tasks require children to recreate the interval they had just experienced using a button press. An advantage of a temporal bisection task was argued to be its suitability for children's attentional capacities (Droit-Volet & Wearden, 2001). However, it seems the authors in this case were comparing bisection of millisecond intervals to reproduction tasks in the range of seconds. Reproduction tasks in the range of seconds have been commonly used for studying developmental disorders such as ADHD (Smith, Taylor, Warner Rogers, Newman, & Rubia, 2002). Difficulties in attending to durations of more than two seconds could explain poorer performance in these individual (Kerns, McInerney, & Wilde, 2001; Luman, Oosterlaan, & Sergeant, 2008; Mullins, Bellgrove, Gill, & Robertson, 2005). On the other hand, reproduction tasks in the range of milliseconds provided a cognitive load that was arguably smaller than temporal bisection because there is no decisional element. Furthermore, reproduction tasks at the millisecond level have been successful in examining children's time perception in recent papers (Karaminis et al., 2016). Considering the suitability of these tasks on the basis of their duration length and associated cognitive load further supports the notion that subsecond and suprasecond timing may be served by different mechanisms.

1.9.2 Modelling the development of time perception.

Bayesian modelling has shown that children's duration estimates may contain larger central tendency biases, due to larger levels of sensory noise. In the to be described study, children showed larger central tendency effects than adults and these effects decreased with age (Karaminis et al., 2016). According to Bayesian inference theories of perception, the purpose of this is to reduce overall error when sensory inputs show high levels of uncertainty. This was supported here because as sensitivity to time increased into adulthood, central tendency effects decreased.

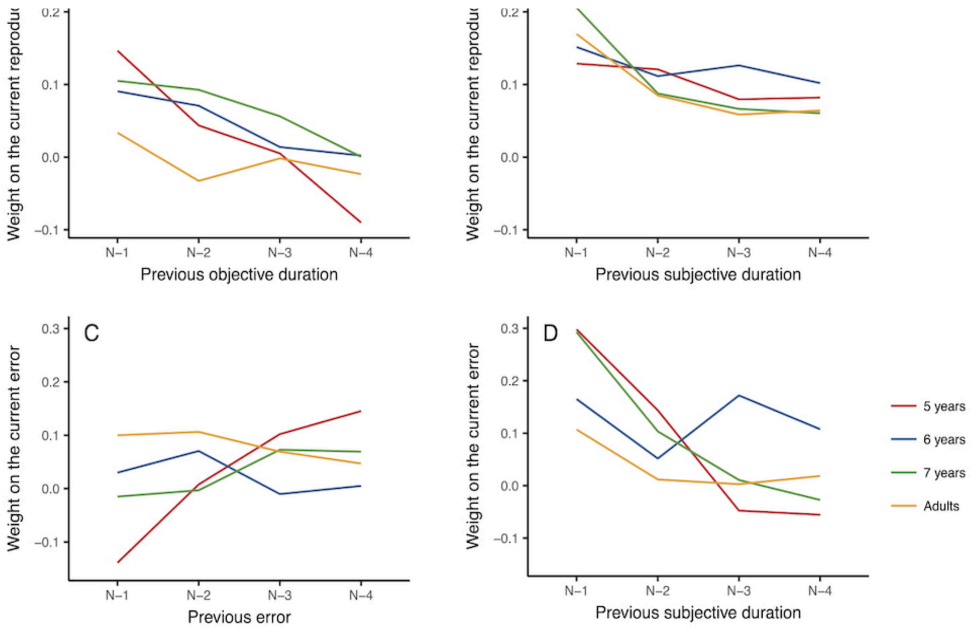
For children with autism spectrum disorder (ASD), their central tendency effects were smaller than expected, given the large amount of variability in their estimates. Their variability levels were more similar to slightly younger typically developing (TD) children (8-9 years) whilst the autistic children were aged between 9 and 14 (M=12). This effect was highlighted due to the ability of the Bayesian model predict the size of the central tendency effects expected, given the amount of variability in the estimates. For children with ASD, it appeared that the central tendency effects did not increase in order to fully compensate for their lowered sensitivity because their central tendency effects were smaller than the model predicted. The build-up of regression in the estimates over the course of the experiment was investigated by doing a split half analysis. In typically developing children and adults, the regression increased from the first to second half. The same was true of the autistic children. Therefore, those with ASD seem to learn a prior at the same rate but there may be differences in how it is acquired. The Bayesian model was able to gain insight into the development of optimal mechanisms for time perception because it incorporated noise (variability) and bias (central tendency effects), and how they varied across different developmental groups.

The mechanism of time perception in children has also been examined using an analytical approach which operated similarly to a multiple regression. Its purpose was to provide an analysis of how well the recent history of task performance predicted the duration reproductions on a given trial. It incorporated the previous objective durations, the previous subjective durations and as predictors. This fixed-factors model used an n-back analysis, where the n refers to the previous trial number, to find that children's reproduction estimates relied heavily upon their previous subjective reproduced durations. This was such that their estimation of the duration that preceded the current trial (N-1) was the largest predictor of their current reproduction length (Hallez et al., 2019). Analyses were also carried out upon the previous six trials with each trial having slightly less influence than the last, as distance from the current duration increased. However, these effects differed across age groups even among children. The youngest children (5 years) showed the strongest reliance upon their previous objective durations with the weight value for N-2 to N-4 diminishing steeply. Whereas older children (7 years of age) did not show a clear pattern of diminishing reliance on the previous objective durations to the same extent (see Figure 12).

Hallez and colleagues (2019) speculated that the reason for the different findings in adults relative to children was immature executive functioning, demonstrated by the children's' reliance on the most recent environmental statistics. This is because they were not yet capable of utilising internal representations to correct errors and reflect on their personal performance (Hallez et al., 2019). Error correction was the strategy used by adults because their current estimate was most heavily predicted by the size of the error they produced for previous durations. Whilst this provided preliminary evidence for children's' use of environmental statistics in order to learn about the world, how their priors may develop differently to adults warrants further study.

Taken together this research gave reason to suggest children rely on perceptual mechanisms, more so than cognitive mechanisms, in early development. This means that their prior formation and subsequent response to timing stimuli may differ from adults because they may have different, and constantly evolving neural obstacles (slow maturing timing sensitivity and executive function) to overcome. Future directions for this area of research could include the study of trial by trial prior development in children. Preliminary findings in this area include Hallez and colleagues (2019) demonstration that children relied so heavily on the previous duration (see Figure 12). A further challenge would be to investigate how this reliance on recent trial history related to Karaminis and colleagues (2016) finding of larger central tendency effects in children (and therefore greater reliance on prior knowledge of the stimulus compared to adults).

Figure 12. The n-back analyses for 5-7 year olds and adults from Hallez and colleagues 2019.



Note. Reprinted from “The dynamic effect of context on interval timing in children and adults”, Q. Hallez and colleagues 2019, *Acta Psychologica*, *192*, pg 91. Copyright *2018 by Elsevier B.V.*

a) The weight on the current reproduction as a function of the previous objective duration

b) The weight placed on the current reproduction plotted against the previous subjective duration

c) The weight placed on the current error against the previous error

d) the weight placed on the current error as a function of subjective duration where the weight in the regression analysis represents the ability of the value on the y axis to predict the value on the x axis.

1.9.3 Autism and time perception

Time perception in autistic individuals has been demonstrated as possessing more variable motor timing (tapping), larger thresholds for duration discrimination and for simultaneity judgements. Additionally, they showed differences in temporal perspective as reported by parents in terms of sense of time and time management (Isaksson et al., 2018). As a result, researchers have proposed that their repetitive and or routine behaviours may present in order to compensate for this poor time sense (Boucher, 2001, as cited by Germanò & Curatolo, 2012) or be related in some way to executive functioning (Allman, 2011). Although with one study using supra-second timing showed equivalent performance in temporal production and estimation, with improved timing performance for reproduction (Wallace & Happé, 2008). In spite of this, there has been a growing consensus that autistic individuals show a large amount of variability in their estimates (Allman & Falter, 2015). This was supported in a recent experiment when children with ASD showed more variability in their responses to free tapping tasks and in their thresholds for simultaneity judgements. Although auditory discrimination thresholds were higher than neurotypical controls (Isaksson et al., 2018). Unanswered questions remain however in relation to both autistic and non-autistic development of time perception. The next section of this essay will explore the spatial domain and Bayesian theories of autism before progressing onto the empirical chapters.

1.9.4 Spatial perception: models and development

Neurotypical development of visuo-spatial abilities

The study of the spatial perception development has used many different paradigms. However, for this discussion of how visual abilities develop, vernier acuity will be a particular area of interest as it is required for successful spatial localisation performance (a task that will be used in Chapter 4 and 5). The development of visuo-spatial processes has been shown through studies with infants, by testing core fundamental aspects of visual perception such as grating acuity (the finest resolvable detail; Skoczenski, 1999) and stereo acuity (the ability to detect depth from the smallest possible angle disparity between objects; Held, Birch, & Gwiazda, 1980). Another type of acuity, vernier acuity, is the ability to distinguish different shapes or stimulus elements (Skoczenski & Norcia, 1999).This type of acuity that is especially relevant to this thesis because the spatial localisation experiment in chapter's 3 and 4, involves detecting and predicting the location of visual stimuli. Therefore, the ability to view the stimuli as separate objects would be necessary for accuracy of perception of the stimuli locations in the spatial localisation task. Vernier acuity is reduced relative to grating (visual) acuity up until around 3 months of age (Shimojo, Birch, Gwiazda, & Held, 1984). It becomes equivalent to grating acuity at 15 months and reaches a level comparable to adults at around 5 years old (Zanker, Mohn, Weber, Zeitler-Driess, & Fahle, 1992). Therefore, development of this ability at 5 years old means that sensing the space between different objects becomes possible and the spatial localisation task can be attempted. Further questions remain about the extent to which young children can predict the locations of different objects during spatial localisation and will be discussed in the following section.

1.9.5 Modelling spatial localisation

Using an adapted version of the spatial localisation task (as described in section 1.5.4) Chambers and colleagues (2018) explored how children's priors may develop in the spatial domain. The difference in the children's version involved the dots that formed the likelihood remained on the screen to account for any working memory or executive functioning differences. After completing the task children under the age of 8 showed evidence of learning the prior mean but not the standard deviation. However, children aged 9 years and older were able to learn the mean and the standard deviation, in a way that was akin to the adults. Therefore, learning about the variability of the environment (and its priors) occurs during development. It will subsequently be important to understand exactly how these mechanisms develop throughout childhood and into adulthood. Future research could work to determine the factors necessary for this ability of prior mean learning to become fully developed by the age of 9 (Chambers, Sokhey, Gaebler-Spira, & Kording, 2018). Furthermore, exploring these questions may reveal other conditions that influence prior formation over a shorter time scale (experiment duration), in addition to the longer time scale of psychological development.

1.9.6 Neurodivergent visuo-spatial development in autistic perception

Research has suggested that autistic people may experience the spatial world differently due to a detailed oriented style of visual perception. Weak Central Coherence (WCC) theory suggests that autistic people have a local, detail-oriented style. Evidence for this idea has been drawn from findings where autistic people have an advantage in processing the local details when typically developing individuals are automatically pulled into global processing (global superiority) and so find it harder to view local features of the stimulus (Happe, 1999; Happé & Frith, 2006).These suggestions were supported in the embedded figures (De Jonge, Kemner, & Van Engeland, 2006; Jarrold, Gilchrist, & Bender, 2005; Pellicano, Gibson, Maybery, Durkin, & Badcock, 2005; Ropar & Mitchell, 2001; Shah & Frith, 1983) and block design tasks (Lockyer & Rutter, 1970; Shah & Frith, 1993; Venter, Lord, & Schopler, 1992). In the case of block design and embedded figures tasks, autistic individuals showed enhanced processing of local detail and less interference from global processing. It then follows logically that if they are less influenced by the top down imposed meaning on the stimuli, they should be less susceptible to illusions that are produced by top down processes. On the contrary, this has not been consistently supported as Ropar and Mitchel (2001) found the performance of autistic individuals on visuospatial tasks did not predict susceptibility to illusions**.** This finding was in contrast with the strong correlation between performance in the visuospatial tasks (embedded figures, block design and Rey test; Ropar & Mitchell, 2001).

In parallel to the above research**,** other researchers focused on enhanced local perception, in the absence of a global perception deficit. Visual search tasks, where a target must be searched for in a grid full of distractors, aimed to understand attention and detail focus in autism. In both feature (where the colour and shape of the target and the distractors were different) and conjunctive search (where the colour of the target and distractors was the same, but the target has a unique shape) autistic participants performed better (Baron-Cohen, O'Riordan, Plaisted, & Driver, 2001; Plaisted, O'Riordan, & Baron-Cohen, 1998). Furthermore, in a triple conjunctive task, where the target was uniquely defined by a combination of 3 features, autistic individuals were better at discriminating targets and distractors than typically developing individuals (O'Riordan & Plaisted, 2001). This evidence led to the development of a theory based on superior discrimination to rival WCC (Mitchell & Ropar, 2004).

This theory was termed the enhanced perceptual functioning (EPF) theory and it explained that autistic individuals experienced superior processing of stimulus elements. However, in contrast with the WCC theory, there is no suggestion that there is a failure of global processing. Support for the enhanced perceptual functioning theory comes from the demonstration of a global advantage, but not global interference, in the Navon figures task (Bellville & Mottron, 1993). The mixed findings surrounding a global motion deficit in autism also make it difficult to conclude that an entire dysfunction of the global processing system exists in autistic perception. Initially increased motion coherence thresholds were seen in autistic groups relative to (verbal age matched) controls (Spencer et al., 2000) and IQ matched controls (Milne et al., 2002). However, other conflicting findings suggested that their motion coherence thresholds were much worse than controls (Pellicano et al., 2005), and Del viva and colleagues (2006) found no significant difference (Del Viva, Igliozzi, Tancredi, & Brizzolara, 2006). Curiously, studies which showed conflicting findings were equally thorough in the process of selecting an autistic (diagnosed) sample.

Other ways of measuring the amount of meaning inferred from global motion has involved the use of optic flow stimuli. This is the visual presentation of stimuli contracting or expanding to produce a subjective feeling of movement. Previous research has shown some posture differences in participants with ASD compared to controls (Gepner, Mestre, Masson, & de Schonen, 1995). Del viva and colleagues (2006) however, found no difference (Del Viva et al., 2006).Similarly, perception of biological motion, the ability to distinguish human motion from point light displays, has also received mixed support. Initially it was found that autistic participants could identify biological motion from point light displays (Moore, Hobson, & Lee, 1997). Although performance on these tasks has showed a negative correlation with autistic severity, implying that there may be some association between biological motion processing and autistic symptoms (Blake, Turner, Smoski, Pozdol, & Stone, 2003). Furthermore, differences in the brain networks activated when completing biological motion task were found, such that there was no superior temporal sulcus (STS) activation in autistic participants (Herrington et al., 2007). Taken together, this evidence suggested that the role of global processing in autism aetiology remains unclear and further investigation is needed to determine the reasons for these discrepant results.

Early research on the Navon task partially explained why a global deficit in autism seemed possible in certain contexts. Navon (1981) found that when participants were presented with a series of large letters, made up of smaller letters called Navon figures and were asked to identify the smaller letters, global interference (the global elements of the stimuli interfere with the global leading to slowed reaction times). Additionally, that global precedence was shown because global elements were processed more rapidly. More recently, however, the Navon task has been presented either in the form of a selective attention task (participants were asked to identify either the global and local stimuli in different blocks) or as a divided attention task (where the target could either be in the local or global features of the stimuli in any one trial). It was found that when autistic participants were given a selective attention task, global interference was shown as it was in TD individuals (Plaisted, Swettenham, & Rees, 1999). The authors concluded that the deficit in autism was therefore likely to be a combination of two factors, the enhanced local processing without overt priming and a difficulty switching from global to local features. This research expanded on the idea of global processing challenges and enhanced local processing because it showed that the specific difficulty highlighted by this task was switching attentional resources. This could explain why embedded figures and block design seemingly shows support for the WCC theory, whilst investigations of global processing deficit in isolation have demonstrated conflicting findings.

An issue with current theories of autistic perception (WCC and EPF) is that they only explain one element of autistic symptoms and whilst there have been attempts to link themselves to social theories such as theory of mind, the exact mechanisms behind such links are unclear (Happé, 1997). Therefore, even if these theories successfully explain some perceptual processes underlying autistic experiences, they do not explain how they come to be this way (Plaisted, 2015), leaving the consequences of complex sensory differences (such as under and over sensitivity to stimuli within the same individual) unexplained (Sapey-Triomphe, 2017).

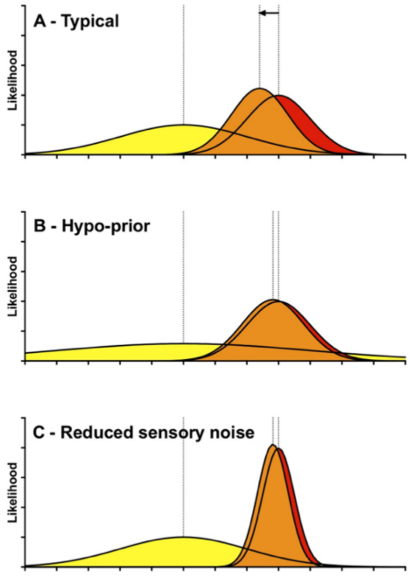
In contrast, the Bayesian framework focuses on the balance of current and prior sensory knowledge. As a result, this theoretical framework could account for a wide range of autistic perceptual differences (Pellicano & Burr, 2012). Initial research suggested that typically developing individuals use prior knowledge to help them make more reliable sensory estimates (by combatting the inherent uncertainty of incoming sensory information; Pellicano and Burr (2012) and this becomes even more essential during unpredictable events. If autistic individuals are less able to weight prior knowledge in these circumstances, this could lead to sensory overload and challenges with unpredictable social environments (Van de Cruys, De-Wit, Evers, Boets, & Wagemans, 2013).

1.10 The Bayesian observer model as a theoretical framework for understanding autistic experience.

Bayesian explanations of autistic experience have focused on a few different aspects of the model: 1) how the prior or likelihood is weighted (Pellicano & Burr (2012), Brock (2013), 2) how priors may be updated with highly precise and inflexible prediction error (Van de Cruys et al., 2013, 2014; Van de Cruys, Vanmarcke, Van de Put, & Wagemans, 2018) and 3) increased learning rate in response to volatility; (Sapey-Triomphe, Weilnhammer, & Wagemans, 2021). Initial theories suggested that autistic people have wider (less precise) priors and that this would be in line with the preference for local detail at the expense of global information described in WCC theory (Pellicano & Burr, 2012). As described in the previous section, the WCC predicted enhanced local processing at the expense of global processing and has been supported through embedded figure tasks (Frith & Happé, 1994), Navon figures (Plaisted et al., 1999) and block design (Shah & Frith, 1993). Bayesian interpretations of autistic perceptual experience have suggested that reduced weighting of Bayesian priors (learning previously perceived statical regularities of sensory input) could be the cause of the difficulties in using previous knowledge in these cognitive tasks (Pellicano & Burr, 2012).

Brock (2012) replied to Pellicano and Burr (2012) and considered that precise likelihoods could be the source and appear with a similar outcome to wide priors (Brock, 2012). This is because weights are given to both the likelihood and the prior where the weight on the prior is w and the weight on the likelihood is 1-w, so a very precise and certain likelihood could also have the same impact on the resulting posterior as a wide prior. This can be seen from the posterior mean remaining the same value in an example where the prior has double the variance (uncertainty) as an example where the sensory likelihood variance has been halved (see Figure 13).

Figure 13. Precise likelihoods or weak priors from Brock (2012).



Note. This figure shows how B and C can result in very similar posterior means (orange).
A) Typical posterior distribution where the arrow indicated the shift of the posterior towards the prior (yellow). This example is a starting point for the variations shown in B and C

B) Here the variance of the prior is doubled (wide hypo-prior) and as a result the shift in the posterior towards the prior is smaller than in A

C) The alternative to B shows sensory noise that has been divided by two. This time the likelihood distribution (red) has become more precise but produced a size in the posterior shift that is almost identical to that shown in B.

Nonetheless, Pellicano and Burr were the first to suggest a Bayesian interpretation to explain that autistic people view the world in a way that is “too real” because they are less able to weight prior knowledge. They argue that their Bayesian interpretation can go beyond the explanations of atypical perception put forward by WCC and EFP by explaining how the mechanisms work computationally. Evidence to support reduced prior weighting comes from Karaminis and colleagues (2016), who found that the autistic children showed reduced bias to the mean compared to the typically developing children who showed similar temporal discrimination performance. The autistic children (9-14 years old) also showed similar amounts of bias to children that were slightly younger than them (8-9 year olds). On the contrary, there is a considerable body of evidence accumulating which appears in conflict with the wide priors (hypo priors) account. This evidence seemed to suggest priors can be used and weighted to some extent in autistic perception. For example, autistic participants were able to follow a visually moving target that would disappear at set intervals with similar predictive eye movements to that of typically developing participants suggesting that sensory predictive abilities were intact (Ego et al., 2016).

Similarly, an experiment, carried out by Van de Cruys and colleagues (2021) provided evidence to suggest that the priors were not less precise in autistic children compared to non-autistic children. They accomplished this through the exploration of contextual (short time scale) and structural (long time scale) priors. In the case of structural priors, both groups demonstrated that they were faster at responding to cardinal orientations than oblique in visual search (Van de Cruys et al., 2021). To study contextual priors, the primed participants with a distractor distribution which meant that for the first few repetitions, there would be a target to find within a group of distractors drawn from a certain distribution. Then on test trials the target would now be presented in orientations that approached the orientations of the distractor distribution. Difficulties arising as the difference between the target and the mean of the distractor distribution demonstrated the formation of a prior because targets now had orientations that were expected from the distractors. It was found from the implementation of Bayesian modelling that there were no differences in the prior variance between autistic and non-autistic children. However, they both showed larger variances than adults when carrying out the same task (Van de Cruys et al., 2021).

As further evidence against autistic individuals showing weaker priors, Sapey Triomphe and colleagues (2021), demonstrated that autistic participants were able to learn strong priors in a time order error task (see Figure 2 for description), but that the challenge was flexibly updating the prior according to a new context. This task involved giving participants a 2AFC magnitude estimation by posing the question which was larger: the first stimulus, or the second? The time order error is a bias where the first stimulus is often overestimated (positive TOE) or underestimated relative to the second (negative TOE; Hellström, 1979). It has been referred to as a contraction bias by Ashourian and Loewenstein (2011) because the perception of the memorised (first) duration becomes biased towards the mean of the distribution. Autistic participants were able to form a strong prior, as evidenced by the production of a visible time order error. However, compared to neurotypical participants, they were less able to adapt their prior as the tactile stimulation experiment changed to include a stimulation range centred on a new frequency (Sapey-Triomphe, 2017). Sapey-Triomphe and colleagues (2021) went on to replicate this finding of prior formation but it was accompanied by a lack of flexible updating. In this experiment participants were asked to estimate the size of a disk as smaller or larger than the interval that followed. The size on any single trial would be taken from either a broad or narrow distribution and when the distribution switched, again participants with ASD did not update their prior to take into account the new distribution width (Sapey-Triomphe, Timmermans, & Wagemans, 2021).

Given the above criticisms of weak priors as a theory, an alternative explanation of autistic prior formation was proposed by Van de Cruys and colleagues (2013). It has been named the Highly Inflexible Precision of Prediction Error (HIPPEA) account (Van de Cruys et al., 2013; 2014). In typically developing individuals when prediction error is used to update their model according to novel but systematic changes in the sensory environment, the precision of that prediction error will be high. However, the precision is lowered when random noise is observed that is, when new learning does not need to occur (Van de Cruys et al., 2014). However, according to HIPPEA the prediction error of autistic people would remain inflexibly high. As a result, model updates occur according to every small change as if it were relevant, even in cases where this would mean learning about noisy, random changes. Priors this precise would lead to trouble with generalisation because it is unlikely that the exact same sensory environment would appear more than once (Van de Cruys et al., 2014).

Another experiment by Sapey-Triomophe and colleagues (2021) provided partial support for this account by finding inflexibility of precision, but not higher precision of prediction error (Sapey-Triomphe et al., 2021). The research was able to demonstrate however that prior acquisition is just as possible for autistic people as for neurotypicals as they both learned the contingency (one sound would play before a dot moved counter clockwise and another tone would play when the dot rotated clockwise), however the contingency was learned to a lesser extent in autistic people and was less flexibly adjusted to the new context when the contingency switched. Other related theoretical accounts have explained that priors in autism are also less able to adjust to the context with no specific neural implementation (Robic et al., 2015) and also that volatility (regular changes to the prior mean or standard deviation) may be a particular challenge (Palmer et al., 2017). Van de Cruys and colleagues (2017) also went on to construct a ‘signal in the noise theory’, whereby it is a difference in the ability to determine which sensory regularities are signal and which are noise as a key component of autistic experience (Van de Cruys, Van der Hallen, & Wagemans, 2017).

Lawson and colleagues (2017) suggested that volatility, the expectation for how causes of sensory input to change over time, was overestimated in autistic adults. A volatile environment contains more frequent changes to sensory regularities. This increased level of surprise often produces an increased learning rate due to a change in expectations about the stability of the environment. Similarly, when an environment becomes less volatile, it is likely that the learning rate decreases again. However, Lawson and colleagues (2017) found that autistic adults overestimate volatility and hence show an increased learning rate. This overestimation may occur as a result of how the external world appears as highly unpredictable to autistic people. Therefore they overestimate volatility and the result is an increase in learning rate to combat this perception (Lawson, Mathys, & Rees, 2017). However, other research conducted in the same year found typical learning rates in response to volatility in autistic children (Manning, Kilner, Neil, Karaminis, & Pellicano, 2017). This raises the question as to whether displays of overestimating volatility (and increased learning rates) could be a compensation strategy developed by autistic adults as they learn to cope with unpredictable sensory information over the duration of their life span. This area of research could also offer insight into the features of autistic experience that become more challenging with age (Happé et al., 2016; Manning et al., 2017). Taken together, this evidence provided evidence for some difference in the ability to adjust to volatile contexts in autistic adults.

The HIPPEA account, and related theories, have the potential to explain a range of differences in autistic perception. The HIPPEA account could explain the origin of autistic experiences such as sensory overload occurring because they treat noisy signals as more relevant than non-autistic individuals do when adjusting prediction error. Similarly, HIPPEA could also explain any discomfort with unpredictability as occurring because learning is less filtered and less flexible. The secondary impacts of are likely to be the elevated levels of stress, anxiety, mood problems and sensory avoidance seen in autistic individuals (Herry et al., 2007; Sapey-Triomphe, 2017). As a result, these accounts appear much more successful in explaining the wide range of differences in autism relative to earlier perceptual accounts such as WCC and EPF.

1.11 Applying a Bayesian framework to individual differences trait research.

As a first step in exploring autistic differences, some researchers have considered that autistic perceptual experience may be viewed as an individual difference at the trait level (Chrysaitis, Jardri, Denève, & Seriès, 2021; Karvelis, Seitz, Lawrie, & Seriès, 2018; Powell, Meredith, McMillin, & Freeman, 2016). The aim of trait research is to explore differences in the general population that relate to the diagnostic criteria of certain conditions either instead of, or in addition to, studying individuals from that population. Relating the degree of, or number of these traits to individuals experience of the condition can be an oversimplification. However, it can also provide an initial exploration of behavioural or perceptual features believed to be a component of the condition and how these features may correlate with the number of traits present in that individual. These studies can, and often do, include individuals diagnosed with the condition.

Karvelis and colleagues (2018) found that autistic traits correlated increased weighting of sensory information. The authors also collected schizotypal personality trait data but found no such connection to likelihood weighting. They used a motion estimation task where participants were required to estimate the direction of the randomly moving dots. To allow the sensory statistics of the stimuli to inform the prior, they increased the frequency of dots appearing at the -32 degrees and 32 degrees directions. This increased the probability of a motion in these directions appearing on screen, resulting in the direction of 32(positive and negative) degrees as the prior. There was evidence of prior acquisition because attractive biases towards 32 degrees were seen in motion direction estimations and estimations of this direction were less variable. It was found that those with high autistic traits demonstrated weaker biases towards the prior direction. Reduced bias under a Bayesian framework could have indicated wider priors or narrower likelihoods. However, because this bias was displayed in conjunction with reduced variability of direction estimation, then narrower likelihoods can be assumed here.

Powell and colleagues (2016) aimed to explore motion estimation priors and autistic traits and considered the flatter prior hypothesis to predict that autistic traits could correlate with reduced weighting given to the prior. The focus was on the visual motion prior of speed where slower speeds are more common and therefore, speeds are often underestimated as a result. More specifically they induced the Aubert-Fleischl phenomenon (where smooth pursuit eye movements decrease the reliability of estimates leading to increased motion sensitivity thresholds and decreased speed estimates) in their first experiment and lowered contrast in their second in order to induce the slower speeds prior in both cases. In Experiment 1, two types of stimuli were designed to produce different types of eye movements (and hence the Aubert-Fleischl phenomenon) during speed discrimination responses. For the fixation stimuli, the annulus window and the fixation remained stationary and only the dots moved (creating the appearance of the dots moving over the window like a sheet). Whereas during the smooth pursuit conditions, the dots, the fixation and the window all moved at the same velocity.

In Experiment 2 participants were also asked to discriminate the stimulus speeds but this time, they were asked which of the horizontal gratings moved more quickly. One stimulus was low contrast and the other was high contrast in order to induce slower motion perceptions, and higher thresholds for the low contrast stimuli. In experiments 1 and 2, the predicted phenomena were demonstrated such that smooth pursuit eye movements produced higher thresholds and slower speed estimates than fixation. Similarly, low contrast stimuli produced higher thresholds than high contrast stimuli (Powell et al., 2016).

The purpose of the model fitting was to answer the question of whether the differences in thresholds could predict the individual differences in the phenomena shown. They compared the fit of their Bayes model to two other models, one with sensory thresholds only and one with autistic traits only. Through this process, Powell and colleagues (2016) found that the Bayes model fit was much more successful in explaining the data than the other models. Additionally, the autistic traits were able to explain a large proportion of the variance in the data. Although the authors postulate that other factors might be influencing the individual differences in these motion phenomena such as how priors were learned and the shape of the prior distributions.

Overall trait research demonstrated key behavioural and perceptual differences according to how many traits the individual possessed. This is useful for initial exploration of these effects or if behavioural tasks are early in development and may need certain modifications before being used with certain populations. It is also important for instances where large numbers of individuals with a diagnosis may not be accessible for the research. Then more participants with measured traits can also be included in correlation based research

From the findings of Glasauer & Shi (2021a), it also seemed possible that participants could have different beliefs about the relationship between stimuli. This is because participants assumed different amounts of serial dependence in the stimuli and therefore demonstrated potential individual differences in generative assumptions when viewing the same duration stimuli (see Chapter 2 for more details). This could be a consequence of how prior distributions are inferred from stimuli and how this drives different models to be selected (Glasauer & Shi, 2021a). Although how and why individuals may vary in the selection of generative models remains to be seen.

1.12 Thesis aim

The overall aim of this thesis was to add to recent explorations of Bayesian prior acquisition in the literature (Glasauer & Shi, 2021b, 2021a; Petzschner & Glasauer, 2011) and to provide additional insights into how priors are updated over time. Another motivation of this research was to see which prior acquisition strategy was most commonly used and whether the prior acquisition process can be flexibly influenced by certain stimulus features. The following chapter (Chapter 2) describes an investigation of how priors are formed during temporal reproduction in relation to different stimulus features. Chapter 3 aimed to answer the question of whether prior formation in the temporal reproduction tasks differed according to autistic traits.

In Chapter 4, the ability to update priors whilst predicting the spatial position of the stimuli was tested via a spatial localisation task. The possibility of switching between, or showing flexibility in, prior updating strategies was also considered. Chapter 5 explored how priors may be updated during spatial localisation throughout childhood and into adulthood. Chapter 6 investigated the possibility of using a Bayesian model to explain biases when comparing two stimuli. As a consequence, Bayesian mechanisms were considered as a potential explanation of a time order error like effect during the reproduction of a single temporal interval. Chapter 7 provides discussion of the conclusions drawn from this research and suggestions for future work.

2. How do priors for interval reproduction update over time: comparing simple iterative, static, and two-state Bayesian models?

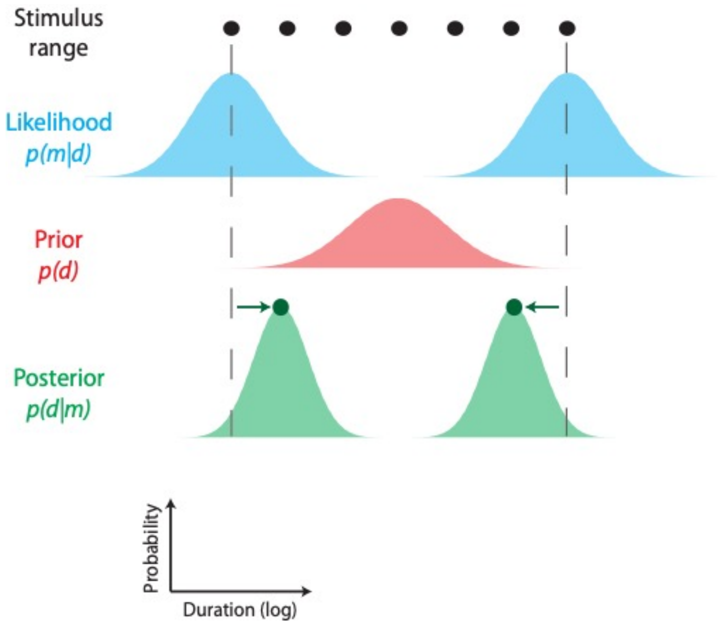
2.1 Introduction

Human sensory estimates are likely influenced by perceptual history due to the rich information provided by accumulated perceptual experience, and the usefulness of this experience for sensory decisions. However, using perceptual history to update our current model of the sensory world may produce a source of bias. Research conducted to understand the principles of successful perception has found that biases were an un-intended consequence of how our perceptual apparatus optimises sensory estimates (Jazayeri & Shadlen, 2010). A potential explanation of these biases could be due to how sensory magnitudes are perceived relative to the context in which they were learned (Petzschner et al., 2015). According to Bayesian accounts, absolute accuracy may be reduced in order to increase the overall reliability of sensory magnitude estimates in a given environment. As a result, sensory estimations may be optimal for the current learning environment (Jazayeri & Shadlen, 2010; Petzschner, 2012)

It has been demonstrated on many occasions that perceptual history has influenced the size of magnitude estimations. The components of previous perceptual experience that may have such an impact include; the modality of the surrounding stimuli (Behar & Bevan, 1961; Goldstone & Lhamon, 1974), the magnitude of surrounding stimuli (Ekman, Frankenhaeuser, Berglund, & Waszak, 1969; Goldstone, Lhamon, & Sechzer, 1978; Lu, Hodges, Zhang, & Zhang, 2009) and the central tendency (mean) of the stimulus distribution (Jazayeri & Shadlen, 2010). It is likely that these changes to the perceived stimulus magnitude were produced as a consequence of optimal Bayesian behaviour that is, giving weight to previous perceptual knowledge. A unified account for the mechanism driving central tendency biases was not available until the suggestion of the Bayesian framework (Jazayeri & Shadlen, 2010; Shi, Church, & Meck, 2013). Under this framework, both current and previous stimulus features are able to influence magnitude estimation. This occurs through the weighted combination of the prior (accumulation of previous sensory input) and likelihood (current sensory input) respectively (see Figure 14).

In Figure 14, the distribution that represents the resulting estimate (posterior) can be seen drifting inwards towards the prior, and away from the raw sensory signals present in the likelihood. Overestimation of short durations and under estimation of long durations occurred because the estimates of the stimulus drifted towards the learned prior (mean average of all physical durations presented previously), producing a biased posterior. In this way, priors act to increase reliability when sensory input signals are uncertain. Therefore, the weight placed on the prior, assuming that the value of the prior is the mean of the stimulus distribution, leads to the central tendency bias in sensory magnitude estimates (Glasauer & Shi, 2019; Petzschner et al., 2015).

Figure 14. A duration stimulus range showing two likelihood distributions, their prior and posteriors.



Note. The likelihoods shown in blue have a peak at each end of the stimulus set. The prior (in red) shows a peak (mean) at the centre of the two likelihood distributions. Due to the posterior being the weighted sum of the likelihood and the prior, this posterior has peaks that are located between their corresponding likelihood peak and the prior.

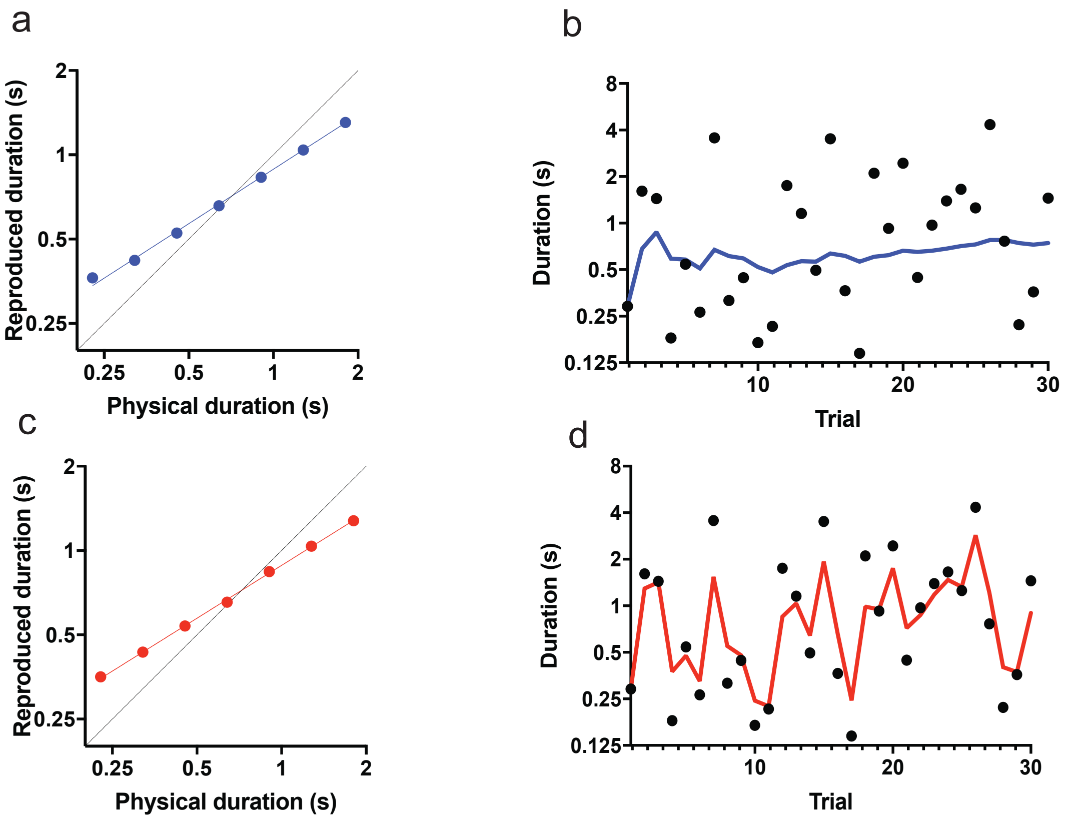
Jazayeri & Shadlen's (2010) research demonstrated that central tendency biases of temporal reproduction could be understood in terms of the Bayesian observer model framework**.** This is because 1) the mean of the stimulus distribution (the prior), was the quantity that future estimates gravitated towards and 2) there was more bias towards the mean as the uncertainty (length of) durations increased. As a result, the prior can aid the inference of the most likely value of a sensory magnitude for any given task, through accessing the previously experienced sensory statistics of the environment. Subsequently, magnitude judgments can be said to be governed by the weight given to the prior and sensory input (likelihood) respectively (Rao et al., 2002). Overall, this leads to a strong influence of the perceptual history on magnitude judgements, especially in contexts containing uncertain or ambiguous sensory inputs, as the prior becomes weighted more heavily under these circumstances (Adams, Kerrigan, & Graf, 2010).

Different Bayesian estimators were explored in Jazayeri and Shadlen's (2010) research to see whether central tendency biases of timing were consistent with the Bayesian ideal observer model. According to this model central tendency biases are produced when a weighted combination of the prior and likelihood (the posterior) is formed, and subsequently mapped onto an estimate. They compared the different Bayesian estimators (rules for mapping) including the maximum likelihood (a uniform prior), the maximum a-posteriori (the mode of the posterior used for mapping) and the Bayes least squares (the mean of the posterior used for mapping) estimators and fitted them to participants responses from a time interval reproduction task. The best fitting model was found to be the Bayes-least-squares estimator (Jazayeri & Shadlen, 2010).

This research was a crucial step in gaining insight into central tendency bias and how it may relate to optimal Bayesian behaviour. Static Bayesian priors, such as those used in Jazayeri and Shadlen (2010), have since been proposed in other Bayesian modelling papers to date (Cicchini et al., 2012; Roach et al., 2017). Such research provides a strong foundation of evidence that priors are weighted according to the reliability of the current sensory signals, however, the crucial next step is to understand how exactly priors are acquired and updated over time.

Understanding the precise nature of how priors update will be essential for improving Bayesian model's ability to explain human behaviour. Multiple models can provide the appearance of a strong fit when looking at summary measures of collated data (the size of central tendency bias) with no consideration of how the prior was updated over the course of the experiment (see Figure 15). The mechanisms that operate within two models that appear to fit equally well on the surface, could be different upon the examination of how their values change over time. As a result, research needs to consider how different generative model assumptions (how the stimuli were generated) can lead to very different methods of updating and forming priors (see Figure 16). Some models may provide more reliable estimates than others due to how sensory information from everyday life presents itself. For example, some events are clearly independent of one another and there would be no reason to think they were temporally related. An example of this would be fully random events (independent and identically distributed). In response to this type of sensory environment, a mechanism for prior updating could involve learning features of the stimuli distribution by approximating the mean and standard deviation.

Figure 15. The possible trial by trial prior updating strategies in central tendency bias data.



Note.

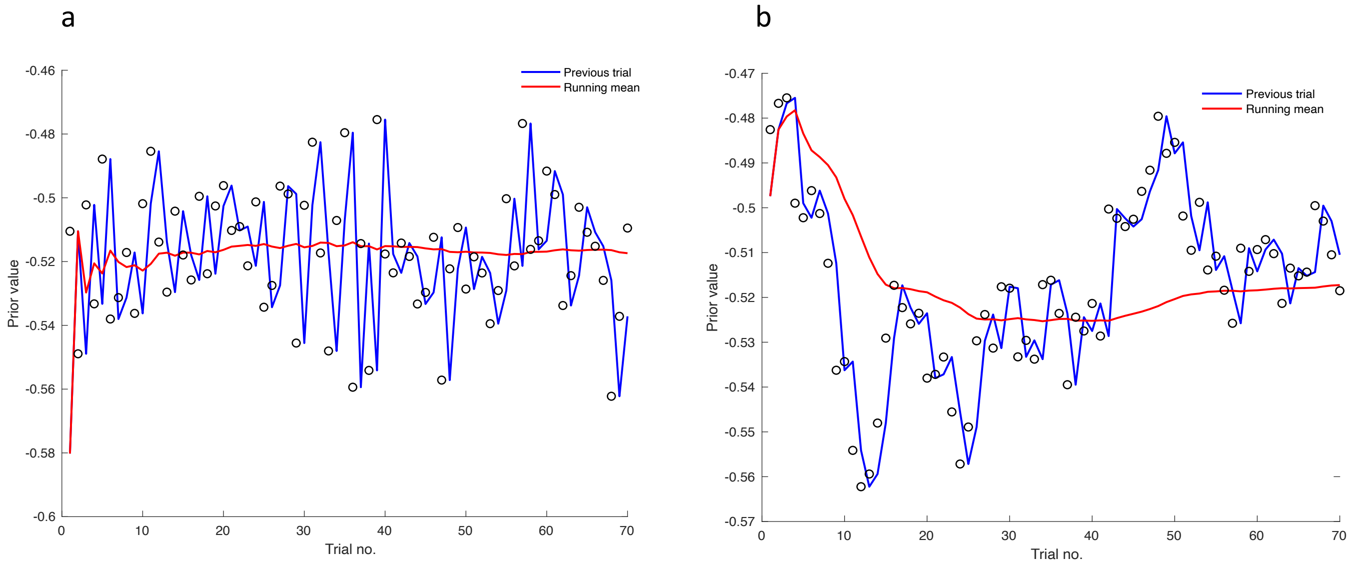
a) The central tendency bias produced when plotting the outcome of reproduction estimates against their physical values. In a) and c) central tendency bias can be seen as an outcome of collecting data, regardless of prior updating strategy. The blue indicates that a static model was used.

b) The mean is approximated across trials and as a result the mean is used to update the prior in a static model.

c) The central tendency bias is the same as that produced in 15a but this time the red line indicates that a simple iterative or Kalman filter model may have been the prior updating process.

d) There is a prior update occurring in response to the value of the previous trial and so the most recent value of the stimulus is used to update the prior.

Figure 16. Bayesian prior updating strategies.



Note. Bayesian prior updating strategies shown through plotting the prior value against trial number. The simple iterative and static models prior updating for (a) random walk stimuli and (b) fully random stimuli. The open circles represent the duration of the stimulus on each trial. The plotted lines represent how each model learns about the value of the stimulus according to their assumptions.

There is evidence to suggest that learning simple features of the distribution (M and SD), along with other more complex features of distribution such as the skewness, bimodality, and kurtosis (tailed-ness of the distribution; Acerbi, Wolpert, & Vijayakumar, 2012) is possible. A generative model which assumes that the stimuli were drawn from a static prior distribution, and learns that distribution, could successfully reduce the variability in sensory estimates in the case of independent and identically distributed events. As a result, learning the characteristics of the distribution would provide the optimal information for prior updating. In addition, there is no need for the mean of the prior to shift according to relevance of more recent or less recent events because the relationship of the events to one another is entirely random.

However, in other cases, there may be reason to think events in the recent past might be related to what you are currently perceiving, and these events can be described as autocorrelated (for example, volume of talking in a restaurant; Date, 2019). Subsequently, how information from a previous event is incorporated into the prior may vary according to our beliefs about how sensory environments appear. Therefore, different manifestations of prior updating strategies may occur in response to different environmental conditions. In this case, the ideal prior updating strategy would involve the prior being updated according to the value of the most recent stimuli, through the action of a simple iterative model (Petzschner & Glasauer, 2011). A prior updating strategy such as this would be most useful if the stimuli contained some degree of serial dependence.

An iterative prior updating mechanism of this kind was explored by the Glasauer group in relation to angle and distance estimation (Petzschner, 2012). Behavioural findings from their study demonstrated that 1) central tendency effects were found in both distance and angle estimation 2) higher sample ranges (a set of long durations) gave rise to larger biases and, 3) that as the distance and angle values increased, so did the standard deviation. The Bayesian modelling process revealed that a simple iterative model (Kalman filter) was more successful in fitting participants data than a static model. As a result, their research has taken initial steps in analysing how Bayesian priors update on a trial by trial basis (Glasauer & Shi, 2018; Petzschner, 2012; Petzschner et al., 2015)

In a temporal reproduction experiment, Glasauer and Shi (2019) varied the order in which the stimuli were presented. In doing so, they manipulated the trial-by-trial dynamics, a potential cue for the prior updating strategies that have been described so far. A random walk protocol was used to present stimuli that varied gradually over time according to a Wiener process (randomly samples values from a distribution). Glasauer and Shi (2019) predicted that during fully random presentation, the sporadic nature of the stimuli order meant that equally large central tendency effects (due to over and under estimation) could arise from both prior updating strategies. However, during random walk conditions, a Kalman filter (iterative) model would predict a decrease in this central tendency bias. This is because the gradual changes in the stimuli would be well suited to its prior updating strategy, and this led to the reduction of the over and underestimation of the stimuli. A reduction in central tendency bias was shown, yet did not disappear completely, for the random walk stimuli (Glasauer

& Shi, 2019). Therefore, the iterative model demonstrated a better fit to participants' responses than a purely static model.

Glasauer and Shi (2021a) subsequently fitted a two state Bayesian model to the empirical data from previous research into distance and angle reproduction (Petzschner & Glasauer 2011) and duration reproduction (Glasauer & Shi, 2019). Despite both static and iterative models predicting the same levels of central tendency for randomly presented intervals, they aimed to separate them based on their predictions of serial dependence. This is because static models predict no serial dependence whatsoever whereas some level of serial dependence is integral to the operations of the simple iterative model (Glasauer & Shi, 2021a). Glasauer and Shi (2021a) presented a two-state model that could incorporate both extremes of the static and the simple iterative model by including two free parameters (v/r and q/r). The v/r parameter defines the amount of variability associated with the prior distribution and the q/r parameter defines how much the mean iterates (shifts) from trial to trial. It can incorporate the possibility of a simple iterative model, where no variability is associated with the distribution (so v/r = 0) and can also account for the possibility of a purely static model (where q/r = 0). The values of each parameter also allow for any possibility in between because a certain degree of variability could be captured by the v/r parameter and a degree of iteration could be captured by the q/r parameter.

By revisiting previous data from the distance and angle reproduction study, they were able to demonstrate that the two-state model provided better predictions of participants serial dependence than either a static or simple iterative model (Kalman filter). The same was true for the duration reproduction experiment and they suspected the two-stage model would explain participants duration reproduction data in both the random walk and fully random conditions (Glasauer & Shi, 2021a). As a result, they predicted the increase in serial dependence for the random walk condition would be accounted for by fitting the same parameter values of the two-state model to both conditions.

Their predictions were confirmed by how successfully the individually fitted parameters from the two-stage model fit in the fully random conditions were able to simulate participants data from the random walk condition. However, for one participant there was no central tendency in the random walk conditions whatsoever, suggesting that participants varied in their beliefs about the amount of serial dependence present in the stimuli (Glasauer & Shi, 2021a). The current research aimed to explore several different prior updating strategies, and which of those strategies best reflected the participants data. Additionally, to see how participants may have demonstrated different best fitting strategies in response to different stimulus conditions.

The first prediction for the current research was the possibility of an iterative model providing the best fit for both stimulus sequences (random walk and fully random). Glasauer and Shi (2019; 2021b) found that a reduction in central tendency bias for random walk stimuli compared to the randomly presented stimuli. This implied the use of a simple iterative model because reduced central tendency would be expected when the stimulus sequence was best suited to the stimulus order (Glasauer & Shi, 2019, 2021b). There is the possibility that the current experiment could replicate this finding of reduced central tendency bias for random walk compared to random stimulus presentation. However, it is also possible that no systematic difference in central tendency bias is found, which could suggest the use of a static model because, in that case, the acquired static distribution does not change according to stimulus order. There is also the possibility that the two-state model provided the best fit for the data using its two parameters to allow for a hybrid of the simple iterative and static models.

In order to explore other contextual cues that could lead participants towards an iterative style of updating (beyond stimulus sequence), the current research also aimed to understand the impact of perceiving stimuli as single and multiple objects.
There is reason to believe, that for spatial quantities (such as location), prediction via a Kalman filter may be used to track a single object for example, the human body during path integration (Petzschner & Glasauer, 2011), and the hand during reaching behaviour (Verstynen & Sabes, 2011). At the current time, little is known about whether viewing the stimulus a single or multiple objects would influence estimates in the temporal domain. Although, if the features of the time interval stimuli signalled the

need for sensory information to be tracked over the course of the experiment, such as appearing as a single object, then it could be most suitable for participants to update the prior gradually according to the most recent stimulus magnitude.

In contrast, if the stimuli appeared more like multiple objects, it could be advantageous for the prior to contain some statistics of the distribution (mean and standard deviation) to summarise the relationship between the objects. The predicted influence of object type was an exploratory hypothesis based upon the findings of Glasauer and Shi (2019; 2021b) in relation to Kalman filters. The current research also used computational modelling to see which prior updating strategy best reflected participants data in response to different stimulus features.

2.2 Experiment 1: Unfilled intervals reproduction task

2.2.1 Methods

Participants

There were 26 (13 female) participants from ages 20 to 43 (M=23.81, SD= 5.48**).** One participant had to be excluded due to outliers (>3 SD’s away from the mean). Normal or corrected to normal vision was a requirement of all participants.

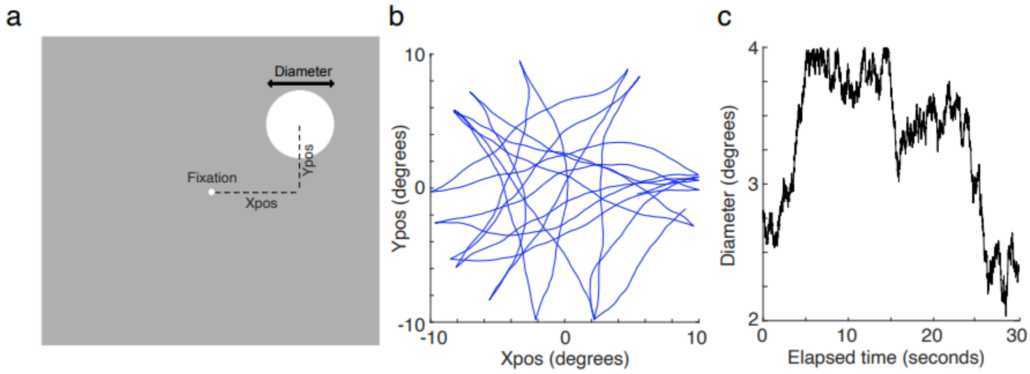
Stimulus details

The stimuli were presented to participants on a Cambridge Research System Display++LCD monitor. The pixel resolution was 1920 x 1080 with a frame rate of 120 HZ. The viewing distance was 62.7cm. At this distance 1 pixel subtended 2 arcmin of visual angle. The display was gamma corrected to ensure linear output.
The target stimulus was a circular disk that moved on a pseudo-random path during inter-trial periods. The speed it moved at during these inter-trial periods was 4 deg/s. At the start of each testing session the stimulus (with an initial diameter of 3 degrees) was randomly positioned within a 20 degrees diameter circular region, centred on fixation and assigned a random direction (see Figure 17a). The direction of the

stimulus was jittered by 5 degrees at the end of each frame, with an equal probability of clockwise or counterclockwise change. If this stimulus path were to reach the edge of the circular region, it was reoriented back towards central fixation (see Figure 17b). In addition, on each frame, the diameter of the target was randomly jittered by 2 arcmin with an equal probability that it could increase or decrease in size. However, this was in the bounds of a minimum diameter of 1 degree and maximum diameter of 4 degrees (see Figure 17c).

When features such as size and position covary gradually within the same stimulus, the stimulus is more likely to be treated as a single object (and learning about that stimulus would be continually updated as opposed to assigning fluctuations to the presence of a new object).Therefore, the size and positions of the target were varied during the inter-stimulus interval in order to provide additional distinctions between each condition (stimulus type). These distinctions between the conditions were displayed clearly by the target remaining on screen during the ISI for single object conditions whereas for the multiple object condition the target disappeared during the ISI. However, The size and position of the target did not change during the presentation of the stimulus interval.

Figure 17. The stimulus size and path for the temporal reproduction task.



Note.

a) The target stimulus was a circular disk randomly positioned within a 20-deg. diameter circular region centred on fixation and a diameter of 3 deg.

(b) The path taken by an example stimulus for a 30 second period. If the stimulus path reached the edge of the 20 degrees circular presentation region, then its direction was reset towards fixation.

(c) The changes in the diameter of an example stimulus occurring over the course of a 30 second period.

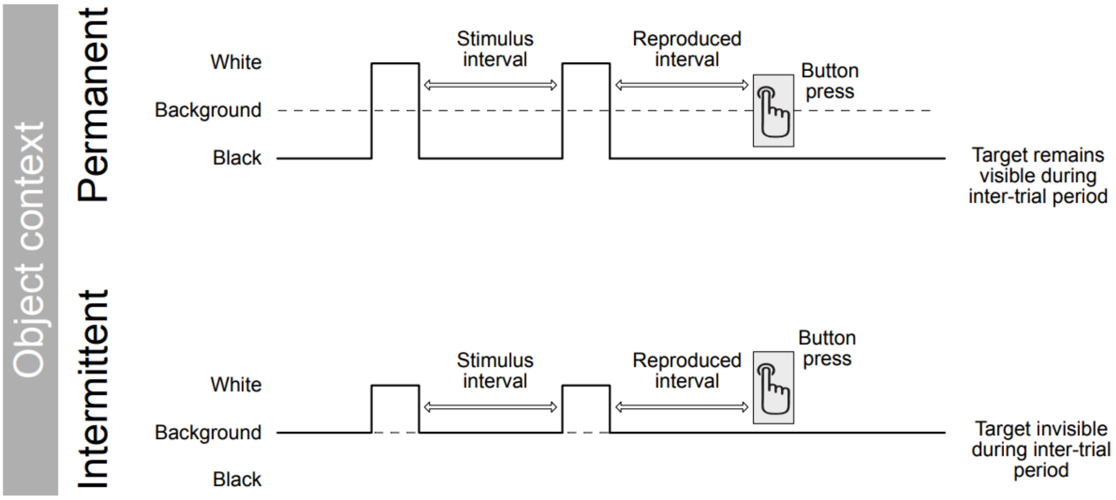
Procedure

Each of the four conditions was ran twice during the experiment to produce four separate testing blocks containing 2 runs each. Then the order of these 4 blocks was randomised for each participant. A single run of the experiment included 70 trials, therefore once each condition had been presented twice, there were 140 trials presented for each condition. As a result, each condition lasted for 5 minutes, leading to a total time of 40 minutes from the start of the experiment to completion but this was often closer to 50 minutes after taking into account breaks.

Participants viewed a white central fixation dot before the first trial began and it remained on the screen for the duration of the experiment. Two stimuli were used to provide the commencement and conclusion of the unfilled time interval. Participants were required to observe the target stimulus (a circular disk) appear on the screen and to notice the inter stimulus duration between the two stimuli. When the white circular disk was no longer visible, they were required to press a button (on a Black Box Toolkit USB response pad).

Participants reproduced the presented intervals using a ready set go reproduction procedure. This procedure, used by Jazayeri and Shadlen (2010), has been named the ready set go procedure because participants must recreate the time interval between the first and second stimuli presentations (ready and set respectively). Participants used their single button press to define where they think a third stimulus would be if the time passed between the first and the second stimulus, and the second stimulus and an imaginary third stimulus, were equivalent (see Figure 18).

Figure 18. Condition details for Experiment 1 (unfilled durations).



Note. In permanent object context conditions, the target was black (<1cd/m2) during inter-trial periods, remaining visible against the mid-grey background (56cd/m2). In intermittent object context conditions, the target had the same luminance as the background in inter-trial periods, rendering it invisible.

During each trial, in experiment 1, the target's luminance was increased in two successive 50ms pulses, demarcating the stimulus interval. Subjects attempted to reproduce the stimulus interval by briefly pressing a button a corresponding time after the second pulse.

Experimental conditions

Two factors that were manipulated within the experiment: object type, which had two levels (intermittent and permanent) and randomisation, also comprising of two levels (full randomisation and random walk). This led to a total of four conditions for this experiment: intermittent and random walk, permanent and random walk, permanent and fully random and finally, intermittent and fully random. As a result, the properties of the target stimulus were varied across different conditions to produce different contextual cues. All participants were exposed to all four conditions in a repeated measures design.

Object type

In conditions where object type was permanent, the target stimulus would appear black in colour (luminance= 1 cd/m^2) during the inter-trial interval. This created the appearance of a target stimulus that was a singular circle alternating from white (trial period) to black (inter-trial period) continually over the duration of the trial run (see Figure 17). The luminance of the white target stimulus was 112 cd/m^2.

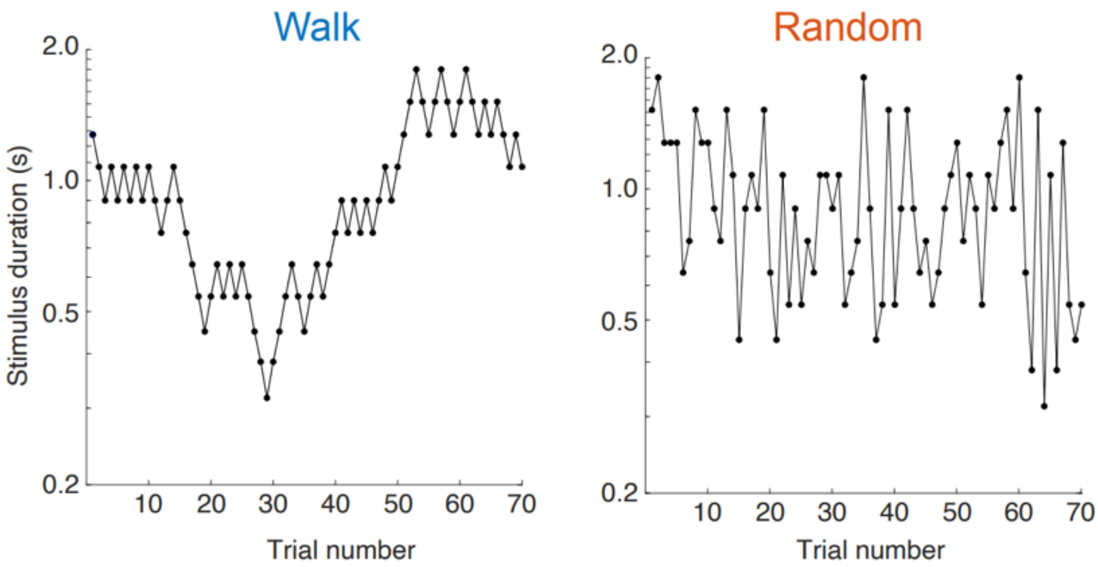
However, when object type was intermittent, the target stimulus was set to be the same colour grey (luminance = 56 cd/m^2) as the background during inter-trial periods. This meant that in between trials, the target stimulus was no longer visible, so it appeared to participants that it arrived at different locations sporadically, even whilst it was still following the gradually moving path detailed in Figure 17b.

Randomisation

In addition to manipulating the object type, the ordering of the durations was also varied. Durations were either presented with full random intervals (shuffled sequence) or in a random walk configuration. For the random walk conditions, the first duration was randomly selected from the 225 ms-1800 ms range. The duration that followed was either increased or decreased by one increment on the 13-point log-spaced scale, relative to the first duration. This one increment change continued for all subsequent trials. The direction of change was randomly determined except for instances where the highest and lowest points of the scale were reached (see Figure 19). To maximise the range of durations that would be observed, the sequence of durations was pre-determined, and this process was repeated if the sequence contained less than 9 of the 13 durations.

For the fully random conditions, durations were again pre-selected from the same set of 13 log-spaced durations (225 ms to 1800 ms) using the random walk procedure. However, in these conditions the order of presentation was randomly shuffled such that there was an equal probability that any of the pre-selected durations were presented on each trial (see Figure 19).

Figure 19. Random walk and fully random stimulus sequences.



Note. Shown here are the sequential ordering of stimulus sequences plotted as a function of duration length. The first duration was selected at random. The next duration in the sequence on each occasion was selected from a possible 13 log-spaced stimulus intervals with an equal chance of a one-step increase or decrease in duration length relative to the duration that preceded it. To ensure adequate coverage of the duration range, the trial sequence was pre-generated at the beginning of each run, and repeated if less than 9 of the 13 durations were included.

For the fully random durations, the predetermined sequence was prepared using the same process as the random walk condition, then these durations were randomly shuffled. This meant that the order of durations was the only factor that varied between the two conditions and not the range of durations presented.

2.2.2 Model details

The Bayesian observer model assumes that the observers compute the posterior probability of each duration given the available evidence. The posterior distribution depends on the likelihood distribution of the current sensory evidence and the prior probability (Petzschner & Glasauer, 2011). According to Bayes rule the mean of the posterior distribution ($\hat{x}_r$) as a generic equation with no prior updating strategy is given by:

$$\hat{x}_r = w_m x_m + w_{prior} x_{prior}$$

Where x_m is the sensory measurement and x_{prior} is mean of the prior. The remaining values, w_m and w_{prior}, are the weights given to those values respectively and the sum of these value is equal to 1. The different Bayesian models fitted to the data included a static model, a simple iterative model and a two-state model (Glasauer & Shi, 2021a). Each model had different assumptions about how the stimuli were generated.

The static Bayesian model assumes that the stimulus on any one trial is drawn from a distribution with a certain mean and variability associated with that distribution. Stimuli are assumed to be drawn from the distribution with a conditional probability of P(x|m) where x is the stimulus and m is the mean of the distribution. The perceptual sensation derived has a conditional property of P(z|x) where z is the sensation of the stimulus, x (see Figure 11).

The mean (peak) of the distribution does not move and is therefore referred to as static. The mean of the posterior distribution (perceptual response) can be written as:

$$y_i = w x_i + (1 - w) x_{prior}$$

Where x is the sensory measurement of the stimulus on the current trial, i, and x_{prior} is the value of the prior mean. The weight, w, is the weight determined by the relative variability of the sensory measurements and the prior.

The simple iterative Bayesian model however, has two central assumptions: 1) that the perception of the stimulus on any trial is the same as the previous trial but with some random noise added to the stimulus 2) the sensory evidence associated with the stimulus is corrupted by measurement noise and for normally distributed noise the Bayesian optimal estimator can be written as a Kalman filter (Glasauer & Shi, 2021a). When the Kalman gain (k) has reached its steady state after a few trials, its equation can be simplified into a weighted average so that the perceptual response (y_i) can be given as:

$$y_i = kx_i + (1 - k)y_{i-1}$$

Where x is the measurement of the stimulus on the current trial, i, and k is the Kalman gain.

Considering the generative assumptions of both the static and simple iterative models is essential in forming the two-state model. For both models described above, measurements are corrupted by sensory noise, so a prior is used in the Bayesian models to reduce the impact of this uncertainty. The way in which the prior is formed depends on the assumed generative model (see Figure 20). For example, for the static model the stimulus, x on trial, i, is assumed to be drawn from the distribution, D(m, v). Where is D is a normal distribution with a mean, m and variance, v. As a result, the generative model (assumption about how the stimuli were generated) can be written as:

$$x_i = m_{i-1} + \varepsilon_x$$

Where εx is a random number coming from a distribution, D(0, v) because the variance, v, when considered relative to sensory noise determines the bounds within this stimulus value can be drawn from.

The iterative model assumes that the stimulus x_i was the same as the stimulus value in trial x_{i-1} but with some random change in variance.

$$x_i = \bar{x}_{i-1} + \varepsilon_m$$

Where ε_m comes from the random shift associated with D (0, q) because q relative to sensory noise determines the amount of intertrial iterative shift.

From these generative assumptions, the two-state model can be constructed (Glasauer & Shi, 2019). The combination of the generative models means that in the two-state model the stimulus value in the current trial depends on: 1) the stimulus value from the previous trial and 2) the shifting mean of the distribution that the current stimulus is assumed to be drawn from. This potential for the stimulus distribution to shift is created because the two-state model assumes that the stimulus x on trial i comes from a random distribution D ($m_i - 1$, v). The mean of this distribution is the same as the mean in the previous distribution but with a random shift in variance q (the same q as described in the iterative model generative assumptions). The two-state model equations are therefore based on the generative assumptions of the iterative and static models (see Figure 11).

The process described in Figure 11 allows the model to now have two simultaneously operating states:

The stimulus being drawn from the distribution:

$$x_i = m_{i-1} + \varepsilon_x$$

and the randomly shifting mean of this stimulus distribution:

$$m_i = m_{i-1} + \varepsilon_m$$

Z, the stimulus corrupted by sensory noise can be described as:

$$Z_i = x_{i-1} + \eta$$

To elaborate on the above equations, on a given trial, i, the stimulus x_i drawn from the distribution with a mean of $mi - 1$ and variance, v. This random number, εx, comes from the distribution N (0, v). The mean of the stimulus distribution is assumed to be the same as the mean on the previous trial with some random fluctuation εm (this is randomly distributed according to N (0, q)). The sensory measurement, Zi, is equal to the actual value of the stimulus, xi and is corrupted by sensory noise, η.

As can be seen from the generative model equations, this two-state model incorporates the variability associated with the distribution and a shifting mean of that distribution on each trial. Therefore, it can incorporate the extremes of each model and any hybrid of the two. This is due to two of its free parameters, the variability of the distribution relative to sensory noise (ratio v/r) and the transition in the mean relative to sensory noise (ratio q/r). The extremes of the static model are made possible by setting the q/r parameter to zero because there is no iteration of the mean on a trial-by-trial basis in the static model. Similarly, the extremes of the simple iterative model are made possible by setting the v/r parameter to zero because the iterative model does not assume that stimuli are drawn from a distribution. The iterative model assumes that the stimulus value on the current trial is equal to that of the previous trial with some random additive shift. The final parameter allows for the upward or downward shift of the model on the y axis, but this is less relevant to the current research.

2.2.3 Results

Empirical Data

Figure 20a shows mean reproduced durations as a function of interval duration for each of the four conditions. For comparison, the dotted diagonal line in each panel indicates the pattern expected for veridical performance. In each condition, group mean reproductions (filled black symbols) show clear evidence of central tendency biases – participants tended to overestimate the shorter durations and underestimate the longer durations. To summarise each participant's performance, we calculated two metrics. First, the relationship between mean reproduced duration ($D_{reproduced}$) and interval duration ($D_{interval}$) was fitted with a power function of the form:

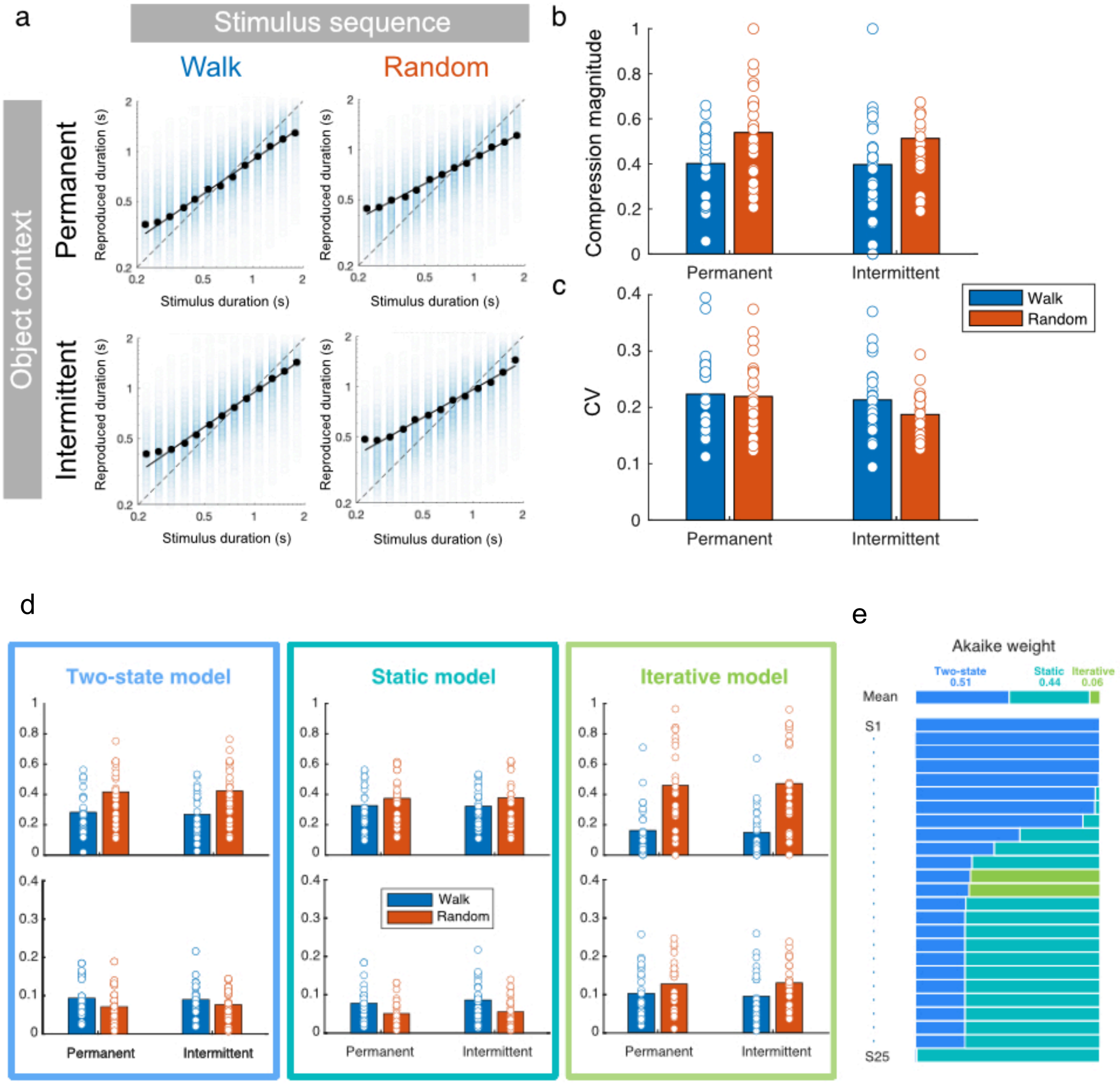
$D_{reproduced} = a \, D_{interval}{}^{b}$

where a and b are free parameters controlling the vertical position and gradient of the power function when plotted on log-log axes respectively. Plotted in this manner, stronger central tendency biases are indicated by a shallower gradient. Accordingly, to capture the strength of this effect, we calculated compression magnitude as one minus the exponent of the best fitting power function:

Compression magnitude = 1-b

Individual and group mean compression magnitudes are shown in Figure 20b. Stronger biases can be seen in conditions with fully random stimulus sequences (marginal mean = 0.589, SE = 0.04) compared to those with random walk sequences (marginal mean = 0.385, SE = 0.03). However, no systematic difference is apparent between permanent and intermittent object conditions. This was confirmed via a 2x2 repeated measures ANOVA, which revealed a significant main effect of stimulus sequence (F (1, 24) = 27.779, $p < 0.01$, eta squared = 0.189), but no significant main effect of object type (F (1, 24) = 0.007, p = 0.934, eta squared < 0.001) or interaction (F (1, 24) = 0.245, p = 0.62, eta squared = 0.003).

Figure 20. Mean reproductions for individual participants for unfilled durations.



Note.

a) Mean reproductions for individual participants (open blue symbols) and the group (filled black symbols), plotted as a function of stimulus duration for each of the four conditions. Solid black lines represent the best-fitting power functions to the group mean data. For comparison, dotted lines indicate veridical performance.

b) Compression magnitudes, reflecting the strength of central tendency biases exhibited by individual participants (open symbols). Bars indicate the mean across participants.

c) Individual and group coefficients of variation, averaged across stimulus duration.
d) The amount of compression magnitude and coefficient of variation shown in model fits for each model.
e) The Akaike weights for all subjects and the mean Akaike weight for each model is shown on the right-hand side.

To summarise the variability of each participant's responses, we then calculated their coefficient of variation (CV_i) for each interval:

$$CV_i = \frac{\sigma_i}{\mu_i}$$

where μ_i and σ_i are the mean and standard deviation of reproductions for that interval respectively. For each participant, a single CV value was then obtained by averaging across intervals. These are plotted in Figure 20c, alongside the group mean values for each condition. A 2x2 repeated measures ANOVA revealed a significant main effect of object type (F (1,24) = 4.808, p = 0.038, eta squared = 0.109), reflecting lower reproduction variability in intermittent object conditions (marginal mean=0.205, SE=0.013) compared to permanent object conditions (marginal mean = 0.224, SE = 0.013). The main effect of randomisation was not significant (F (1, 24) = 2.841, p = 0.105, eta squared = 0.024), but there was a significant interaction between object type and stimulus sequence (F (1, 24) = 5.120, p=0.033, eta squared = 0.021). An analysis of simple effects indicates that this reflects the fact that in intermittent object conditions, reproductions were less variable with random sequences (M = 0.196) than walk sequences (M=0.213, F(1,24)=6.921, p=0.015), whereas there was no significant effect of stimulus sequence in permanent object conditions (random M=0.224, walk M=0.224, F(1,24)=0.004, p=0.950).

Computational modelling - fixed strategy

To investigate the form of prior acquisition strategies employed by participants, we first fitted each of the three models simultaneously (using the 'fmincon' function in MATLAB) to each individual's data from all conditions and calculated performance metrics using the same process described above for the experimental data. Best-fitting models were those that minimised the squared residual error between predicted and reproduced durations on each trial. Summary results for the best-fitting models are shown in Figure 20d, plotted using the same conventions as the experimental data in Figure 20b and 20c.

The iterative model assumes that stimuli on successive trials differ only by some small random fluctuation. This assumption is more appropriate for walk sequences than random sequences, so it is reasonable to expect the performance of the model to vary accordingly between these conditions. This prediction is borne out in the model output – reproductions generated by the iterative model exhibit less compression bias for walk sequences than random ones. This echoes the pattern seen in the experimental data (Figure 20b), although the difference in compression magnitude for the iterative model is noticeably more pronounced. Reproductions from the iterative model also exhibit less variability for walk compared to randomised sequences, again consistent with the relative suitability of the model to these two scenarios. However, it is important to reiterate that there was no evidence for a precision advantage for walk sequences in the experimental data (see Figure 20c).

In contrast to the iterative model, the static model assumes that stimuli on each trial are sampled from a distribution with fixed properties, an assumption that better matches the statistics of the random sequences than that of the walk sequences. This is reflected in the coefficients of variation of the modelled reproductions, which are consistently smaller for the random sequence than the walk sequence conditions. Note, this pattern was also seen in the experimental data, albeit only for intermittent object conditions (see Figure 20c). Interestingly, the static model also captured some of the attenuation of central tendency biases with walk sequences seen in the experimental data. This was a somewhat unexpected result, given that the assumptions of the static model are less compatible with walk sequences than random sequences. Reproductions from the best-fitting static model also showed a slight difference between the amount of mean compression in the fully random and random walk conditions (see Figure 20c).

The two-state model assumes a more complex generative model in which stimuli are drawn from a stimulus distribution, the mean of which changes iteratively from trial to trial (see Section 1.8). Depending on the choice of parameters therefore, it can operate identically to either the static (when q/r is zero) or iterative model (when v/r is zero) alone, or as a hybrid of the two (when q/r and v/r are both non-zero). As shown in Figure 20d, the best fitting two-state model successfully captured the difference in compression between random walk and the fully random conditions. However, like the static model, it produced estimates with reduced variability for random stimulus sequences compared to walk sequences.

To directly compare the three models, we used the sum of squared residual errors (SS_{err}) to calculate the Akaike information criterion (AIC) for each model:

$$AIC_i = N \log\left(\frac{SS_{err}}{N}\right) + 2K$$

where N is the number of reproduced durations (trials) contributing to the fitting procedure and K is the number of free parameters (two for static and iterative models and three for the 2-state model).

We then calculated the relative likelihood of each model:

$$Relative\ likelihood = e^{-\Delta \mathrm{AIC}/2}$$

where ΔAIC is the difference between the model's AIC and the minimum AIC obtained (ΔAIC$_i$ - min(ΔAIC$_i$).

Akaike weights were obtained by normalising each model's relative likelihood by the sum of the relative likelihoods across all three models:

$$w_i(AIC) = \frac{e^{-\Delta_i \mathrm{AIC}/2}}{\sum_{k=1}^{K} e^{-\Delta_k \mathrm{AIC}/2}}$$

These weights can be interpreted as the probability that each model is the best model, given the data and the set of models considered (Wagenmakers & Farrell, 2004).

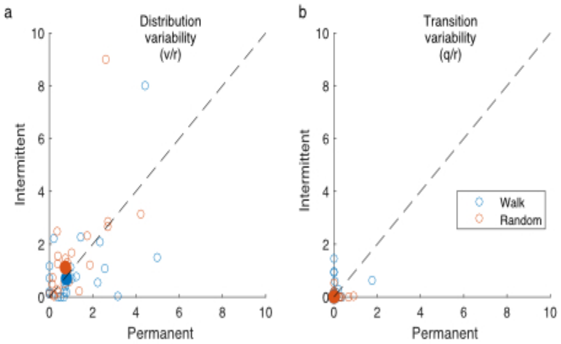
Figure 22 shows the set of Akaike weights for each observer, as well as the mean of the group. Although the Iterative model was favoured for two participants, it generally received very low weighting (mean Akaike weight = 0.06), indicating that it was the least successful as an account of participants' performance across all conditions. The static model received the highest weighting in the largest number of participants (n=14, mean Akaike weight = 0.44), whereas the two-state model received the highest mean weighting (n=9, mean Akaike weight = 0.51).

It is important to note here that the two-state model provided the best raw fits (i.e. lowest mean squared error) for all participants. This is to be expected – it is a higher-order model that encompasses both static and iterative approaches as well as a continuum of hybrid approaches in between them. However, the results of the AIC analysis indicate that for most participants, the benefit of this additional complexity was not sufficient to overcome the penalty applied for its additional free parameter. If we assume therefore, that participants employ a fixed strategy for all conditions, this appears to be best approximated by the static model, or at least a two-state model that is closer to the static end of the continuum than the iterative end.

Computational modelling – flexible strategy

We next considered the possibility that participants switch their strategy depending on the experimental conditions. To do this, we first re-fitted the two-state model to each participant's data, but this time allowed the parameters to vary between each of the four conditions tested. Two parameters control the model's assumptions regarding how stimuli have been generated – distribution variability (v/r, the standard deviation of the stimulus distribution relative to sensory noise) and transition variability (q/r, the standard deviation of trial-by-trial shifts relative to sensory noise). These parameters accommodate for a model where stimuli are assumed to be drawn from a distribution and some iterative shift of this distribution. Therefore, it is possible that different intermediary steps between the two extremes were being used by participants for each stimulus condition.

Figure 21 Best-fitting parameters of the two-state model when separately fitted to each experimental condition (unfilled intervals).



Note.

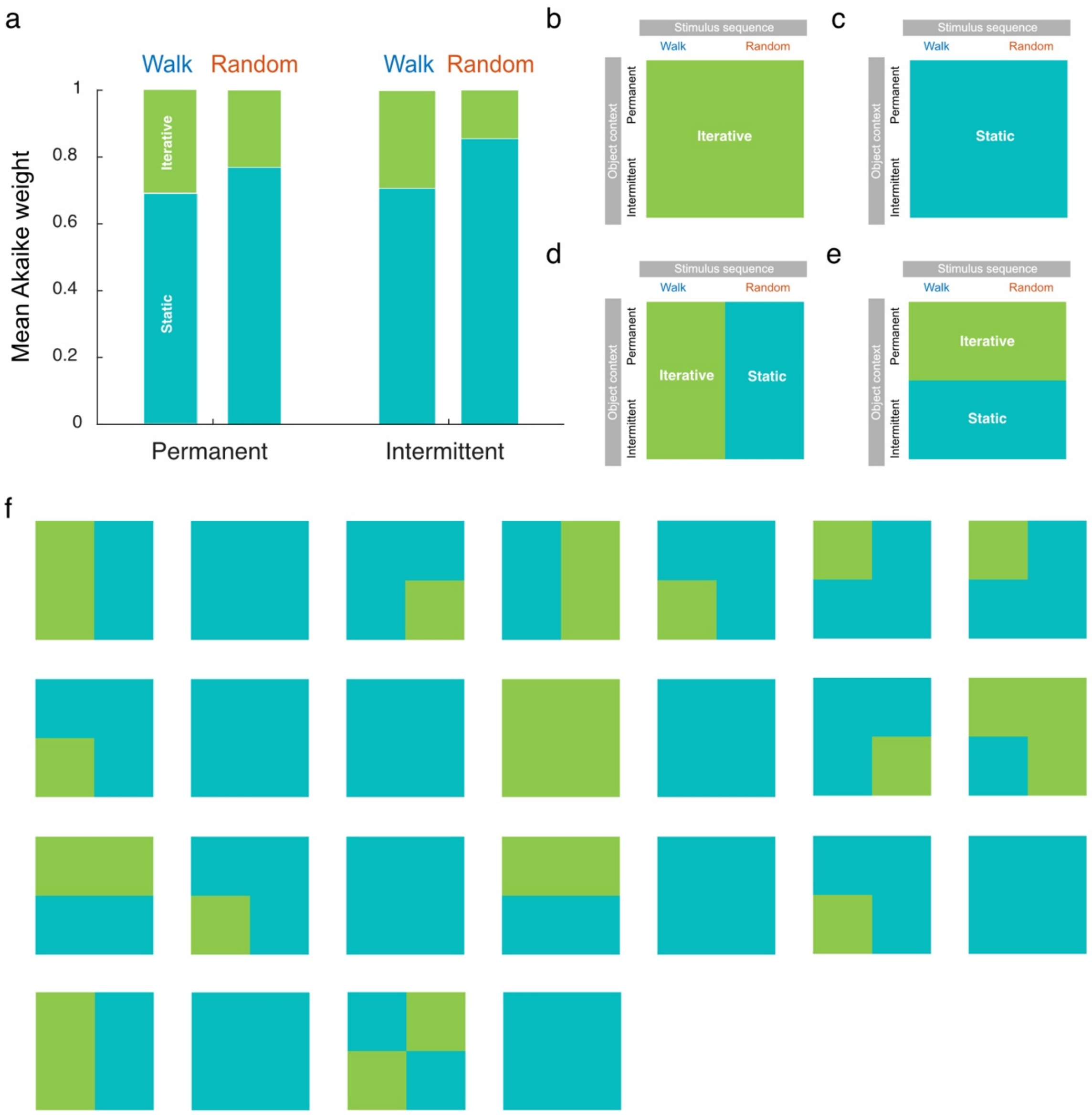
(a) v/r represents the variability of the distribution relative to sensory noise while (b) q/r represents the variability of random additive intertrial shifts relative to sensory noise. Unfilled symbols show results for individual participants, whereas filled symbols denote the group mean.

Figure 21a and 21b show the best-fitting values for v/r and q/r respectively. In each panel, parameter values for participants' intermittent object conditions are plotted as a function of values in permanent object conditions, with results from walk and random conditions shown in blue and orange respectively. Overall, there is little evidence for a systematic change in parameter values between conditions. While there is some variability between fits, distribution variability (v/r) remained greater than zero in the majority of cases, whereas transition variability (q/r) estimates remained clustered around zero. Therefore, participants continued to assume that stimuli are sampled from a static distribution, regardless of experimental condition.

As a final test for flexibility in prior acquisition strategy, we carried out a binary comparison of static and iterative models when separately fitted to each condition. Figure 20e shows Akaike weights averaged across all participants for each condition. Consistent with previous findings, the static model consistently outperformed the iterative model across all conditions. There is a slight trend towards the iterative model faring better for walk sequences than random ones – consistent with a change in strategy that matches the assumed generative model to the stimulus sequence. However, these differences are relatively modest when viewed at group level.

To examine whether these results could be driven by a sub-set of individuals, we next summarised which of these two models performed best for each participant in each condition. These are colour coded in Figure 24f**,** where green and blue indicate superior performance of the iterative model and static model respectively. For comparison, examples of expected patterns under different prior acquisition strategies are represented here as rigid adherence to an iterative model (see Figure 24b), rigid adherence to a static model (see Figure 24c), switching model based on stimulus sequence (see Figure 24d), and switching model based on object context (Figure 24e). Results for 9 participants matched rigid adherence to the static model, with a further 8 participants favouring the static model for 3 of the 4 conditions. In contrast, only 2 participants demonstrated a best fit for iterative model in 3 conditions or more. Although the iterative model was demonstrated to be the best fit for at least 1 condition out of 4 for 16 participants. Two participants showed best fits in line with predictions of stimulus order conditions however, another participant showed the reverse of this pattern. For the object type prediction, 2 participants did show best fits in line with the predictions made despite this not showing a main effect in the ANOVA.

Figure 22. Akaike weights showing the best fit for iterative or static models (unfilled intervals).



Note.

a) The most likely model to be used by participants in each condition according to their Akaike weight in each condition.

b) to e) the predictions of the most suitable strategy for each condition to compare to the actual strategy participants used.

f) All 25 participants according to the most likely strategy used by them in each condition.

2.2.4 Discussion

The aims of this study were to examine the bias in estimates of temporal reproduction in response to different stimulus features and to subsequently determine best fitting models. The compression magnitude data appeared to show some differences between the conditions, through a significant main effect of randomisation. This meant that the size of the central tendency bias was larger for fully random conditions compared to random walk conditions. In this way, the compression magnitude data showed the closest alignment with predictions of the two-state model because it showed some reduction in compression for random walk conditions. The iterative also shows a reduction in compression according to randomisation but was much larger than participant responses.

On the other hand, the variability (CV) of participants responses demonstrated a significant interaction. This meant that participants showed significantly reduced variability in fully random compared to random walk for intermittent conditions only. This reduction in variability for fully random conditions aligns more closely with the predictions of the static model. The static model assumption that stimuli are drawn from a distribution means that variability can be reduced more successfully for fully random conditions.

When exploring the possibility of a fixed strategy, the best fit to the empirical data appeared to be the two-state (closely followed by the static model) whereas the iterative model seemed to provide the least successful fit. Furthermore, as stimulus conditions were manipulated to provide advantageous performance for an iterative model, for example presenting the stimulus in a random walk order, participants did not change their strategy to an iterative model. When allowing the parameters of the two-state model to vary according to condition, there was little evidence of a flexible strategy according to conditions and, in all conditions the transition parameter used to account for iterative processes within the static model was almost at 0. Additionally, participants data continued to show a closer fit to the static model when either an iterative or static model was fitted to each condition.

Overall, this means that a static model or a predominantly static two state model would be most able to account for participants data. However, this is in stark contrast with recent findings of participants favouring an iterative model (Glasauer & Shi, 2019, 2021b). One potential reason for this could have been the unfilled stimuli used to present the duration. Filled durations may be treated differently such that they could be more cognitively demanding due to appearing longer than unfilled durations (Block, Hancock, & Zakay, 2010). Glasauer and Shi (2019; 2021b) used a filled interval method of responding, whereas the current experiment used an unfilled interval. As a result, the current experiment was repeated with filled interval reproduction. All methodological processes were identical other than the response procedure described below:

2.3 Experiment 2: Filled interval reproduction.

2.3.1 Methods

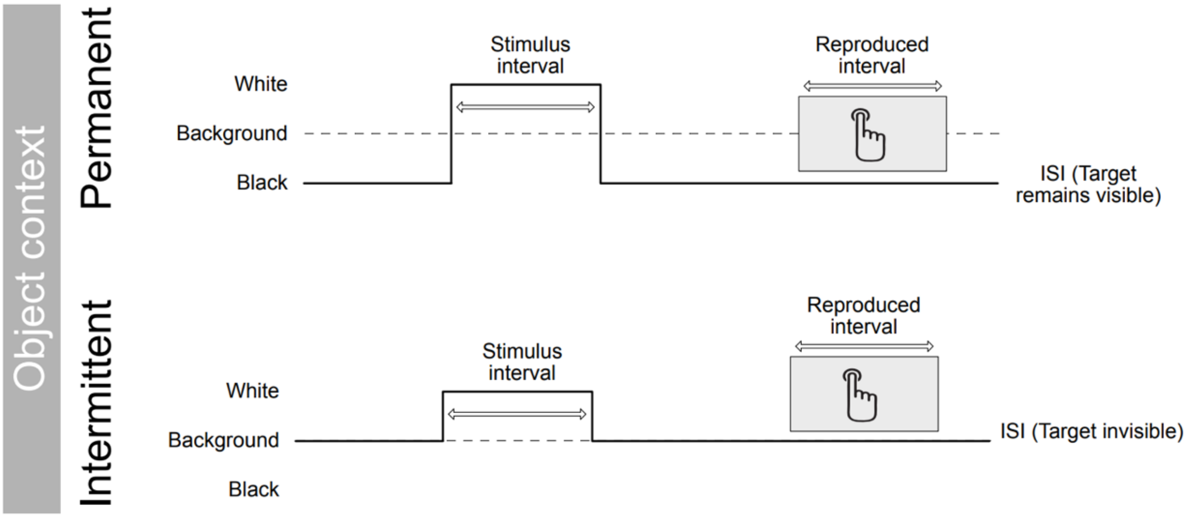
Participants

The filled experiment was completed by 51 participants (2 were excluded due to anomalies more than 2 SD away from the mean or incomplete data sets). The remaining participants had an age range from 18 to 53. All but 4 participants were students at the University of Nottingham. As before participants were required to have normal or corrected to normal vision.

Procedure

In Experiment 2, all elements of the procedure were identical excluding the change from unfilled to filled stimuli. This change involved presenting a single stimulus, which remained on screen to produce a filled time interval and participants responded by pressing down for a period they felt was equal to the time they had viewed the target stimulus for (see Figure 23).

Figure 23. *Condition details for Experiment 2 (filled durations).*



Note. The target's luminance was increased once on each trial, for a duration defining the stimulus interval. Participants attempted to reproduce the interval by pressing and holding a button for a corresponding period of time. All other methods were identical to Experiment 1.

2.3.2 Results

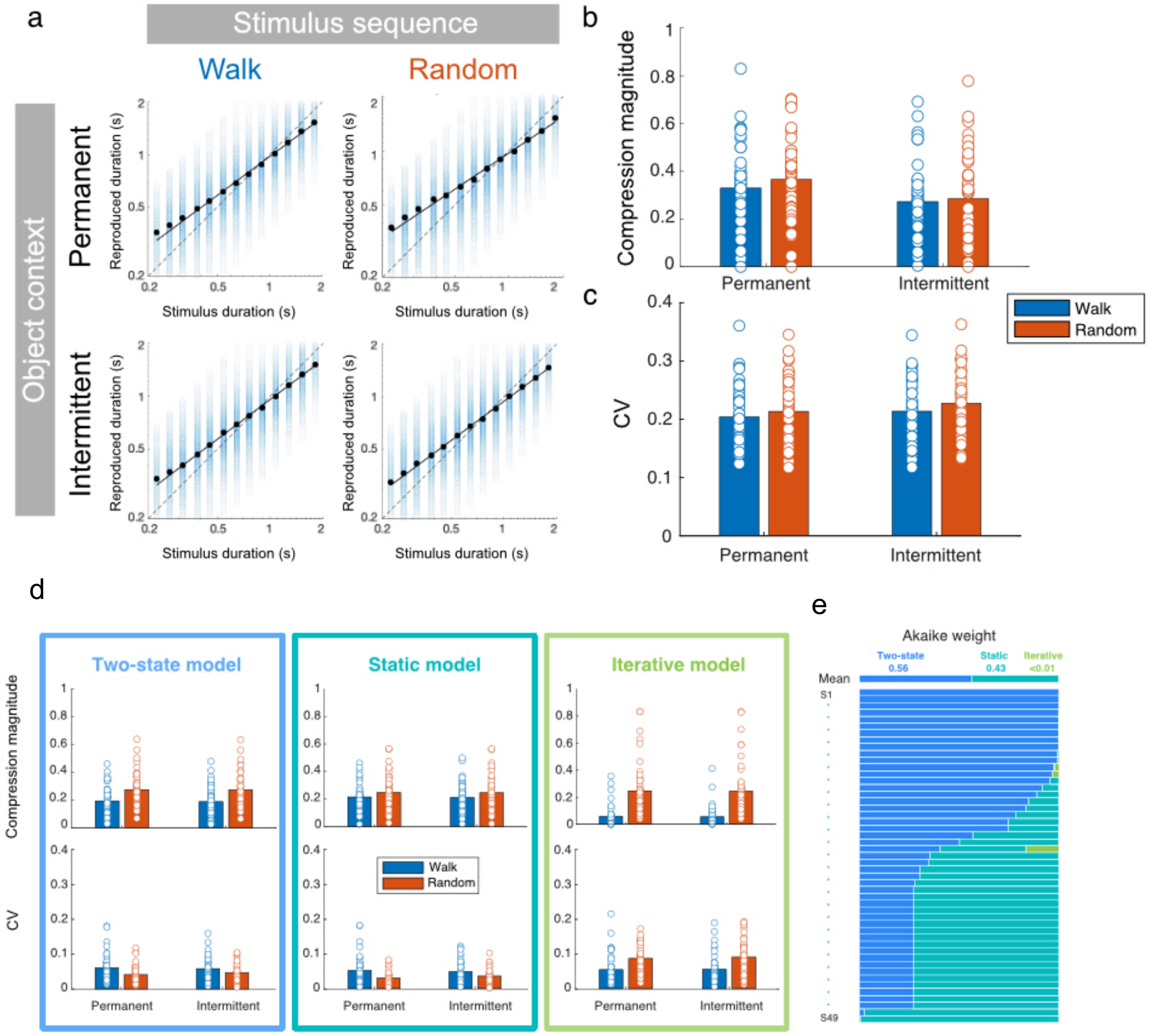
Empirical data

Figure 24a shows mean reproduced duration length as a function of physical duration length for each of the four conditions for the filled duration stimuli. As before, the dotted diagonal line in each panel indicating the pattern expected for veridical performance. The participant performance metrics were calculated again, and the power function fitting process described in the results section of Experiment 1 was repeated to allow for comparison of the compression magnitude across the four conditions.

For the filled durations there appeared to be a very similar amount of compression across all conditions (see Figure 24a and 24b). The bar heights in all conditions showed few differences between one another. There was a very small increase in compression in the fully random compared to the random walk and a more noticeable increase in compression for the permanent conditions compared to the intermittent conditions.

d mean compression across participants showed values between 0.24 and 0.4 in all conditions for the filled duration stimuli (see Figure 24b), therefore displaying smaller compression magnitudes than those shown in response to unfilled durations. For the unfilled stimuli, every condition showed an average compression magnitude across participants of greater than 0.35.

Figure 24.Mean reproductions for individual participants for filled durations.



Note.

(a) Mean reproductions for individual participants (open blue symbols) and the group (filled black symbols), plotted as a function of stimulus duration for each of the four conditions. Solid black lines represent the best-fitting power functions to the group mean data. For comparison, dotted lines indicate veridical performance. b) Compression magnitudes, reflecting the strength of central tendency biases exhibited by individual participants (open symbols). Bars indicate the mean across participants. (c) Individual and group coefficients of variation, averaged across stimulus duration. (d) The amount of compression magnitude and coefficient of variation shown in model fits for each model. (e) The Akaike weights for all subjects and the mean Akaike weight for each model is shown on the right-hand side.

To investigate whether these differences in compression were statistically significant a 2x2 repeated measures ANOVA was conducted. There was a main effect of object type F (1, 50) = 9.646, p =0.003, eta squared = 0.063 and therefore that noticeable difference in compression magnitude between permanent (marginal mean = 0.316, SE = 0.019) and intermittent (marginal mean = 0.258, SE = 0.019) object type was statistically significant. However, there was no significant main effect of randomisation F (1, 50) = 1.491, p -0.228, eta squared = 0.010 and no interaction between the two variables F (1, 50) = 0.054, p = 0.818, eta squared = 0.0003.

The coefficients of variation were calculated as described in the unfilled duration results section. There did not appear to be any graphically visible differences in CV across the conditions. However, a 2x2 repeated measures ANOVA found a main effect of randomisation F (1, 50) = 5.323, p = 0.025, eta squared = 0.023, but no main effect of object type F (1, 50) = 1.815, p = 0.184, eta squared = 0.018 and no interaction F (1, 50) = 0.468, p=0.497, eta squared = 0.03.

In contrast to the unfilled intervals, there appeared to be a larger amount of variability associated with the fully random conditions (marginal mean = 0.221 SE = 0.007) compared to the random walk conditions (marginal mean = 0.210, SE =0.006).

Modelling - fixed strategy

As before, all 3 models were fitted simultaneously, and the performance of those models can be seen in Figure 24d. The features of the models, as shown by the best fits, appear very similar to the unfilled data and this is likely due to how, across the two experiments, the stimulus presentation was kept as similar as possible even as the response type was manipulated.

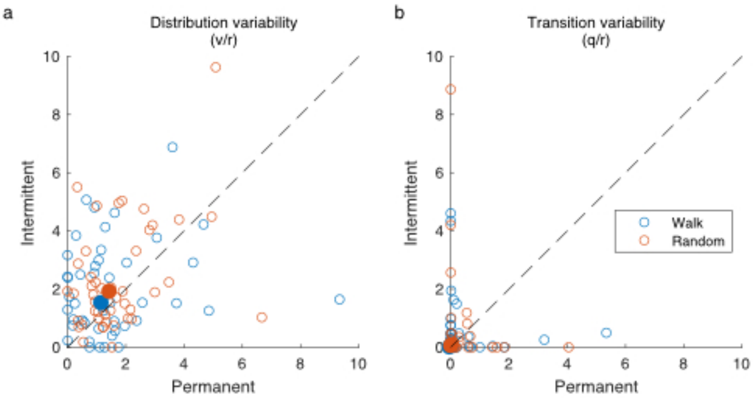
To explore how successfully each model fitted the data, Akaike weights were calculated (as described in Experiment 1). For many individual participants, the blue colour representing the two-state model and the turquoise colour representing the static model were highly prominent (see Figure 24e). For several participants static and iterative were so unlikely that they were not visible in this figure. It is important to note that the Akaike weight penalises the two-state model for having an extra parameter and it continued to be the most likely model used for many participants after being penalised. There were also many participants where the static model shows the highest Akaike weight and therefore the static model was the most likely in their case. However, the mean Akaike weight across participants indicated that the best fitting model overall was the two-state model (0.56), very closely followed by the static model (0.43). The Akaike weight for the iterative model indicated that its use was highly unlikely (<0.01).

Modelling - flexible strategy.

We again considered the possibility that participants switch their strategy depending on the experimental conditions. As before, we first re-fitted the two-state model to each participant's data and allowed the parameters to vary between each of the four conditions tested and the same graphical conventions applied.

Figure 25a and 25b show the best-fitting values for v/r and q/r respectively and are plotted in the same way as the unfilled duration fits. Similarly, to the case of the unfilled durations, distribution variability (v/r) remains greater than zero in the majority of cases, whereas transition variability (q/r) estimates remained clustered around zero. Participants appear to favour a model that assumes a static distribution, even in experimental conditions that may have been better suited to an iterative model.

Figure 25. Best-fitting parameters of the two-state model when separately fitted to each experimental condition (filled intervals).



Note.

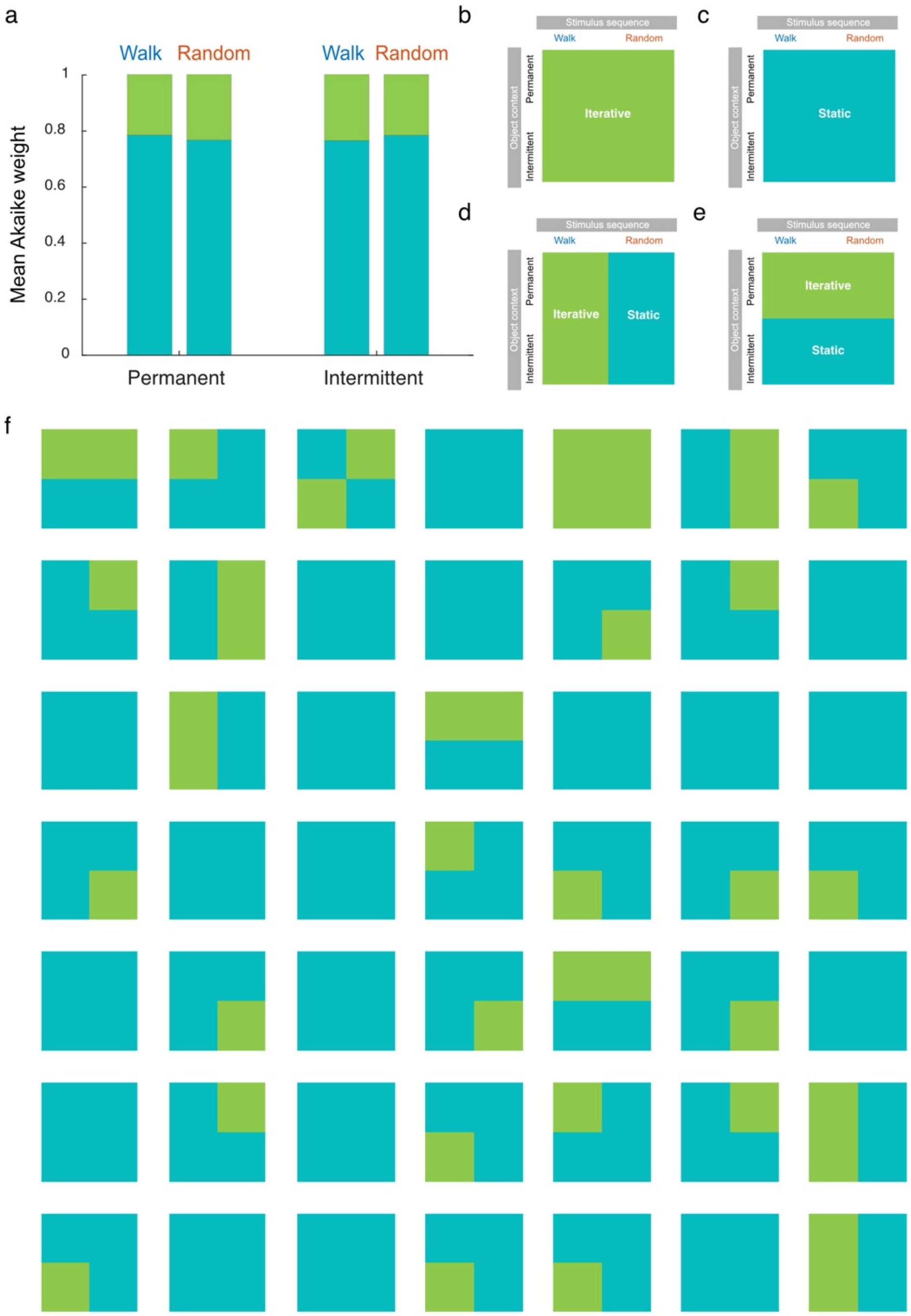
a) v/r represents the variability of the distribution relative to sensory noise while b) q/r represents the variability of random additive intertrial shifts relative to sensory noise. Unfilled symbols show results for individual participants, whereas filled symbols denote the group mean.

As before, to further investigate the possibility of a flexible model, we then compared static and iterative models when separately fitted to each condition. The best fitting model when Akaike weights were averaged across participants was the two state, closely followed the static model (see Figure 24e).

The average Akaike weights for each participant was averaged to create a mean Akaike weight for each model in every condition. There was a very small increase in weight for iterative model (green), in the case of the fully random conditions for the permanent object type. However, there was also a small increase in weight for the iterative model (green) in the case of random walk conditions for the intermittent conditions.

The predictions of the best fitting model according to condition are shown in Figure 26b to 26e labelled as per the previous section (see also Figure 22b to 22e and description of predictions in previous results section). Keeping these predictions in mind, the best fitting model for individual participants in each condition is shown in Figure 26f. For nineteen participants the static provided the best fit across all conditions.

Figure 26. Akaike weights showing the best fit for iterative or static models (filled intervals).



Note.

a) The most likely model to be used by participants in each condition according to their Akaike weight in each condition.

b) to e) the predictions of the most suitable strategy for each condition to compare to the actual strategy participants used.

f) 49 participants according to the most likely strategy used by them in each condition.

For three participants, their best fitting models were found to be in line with predictions for full vs walk that is, they showed a best fit for the iterative model in the random walk conditions and a best fit for the static model in the fully random conditions. Although two participants showed the reverse pattern. Two participants also showed best fits in line with object type prediction. More specifically, their best fitting model was the iterative model for permanent conditions and the static model for the intermittent conditions. The number of participants that displayed model fits in line with the predictions were very small compared to the large number of participants where most, if not all conditions, showed the static model to be the best fitting model.

2.3.3 Discussion

The compression magnitude data shown in Experiment 2 can be accounted for by the static model because the levels of compression in all conditions appear highly similar. Unlike the unfilled durations, the prediction according to stimulus sequence did not produce a significant main effect. Moreover, the prediction according to object type occurred as a main effect in the opposite direction to that of the prediction of the iterative model. As a result, there is not enough evidence to conclude that viewing temporal stimuli as single objects makes the use of an iterative model any more likely.

The variability (as measured by coefficients of variation) in participants responses showed relatively small differences between conditions but there was a significant main effect of randomisation. When the coefficients of variation were plotted, there was a visible decrease in variability for the random walk compared to the fully random which would be in line with an iterative model. However, the size of this difference was not as large as that predicted by the iterative model. The static and two state models would predict smaller differences in variability between the two different stimulus sequence conditions. Although they also predict that, contrary to the current pattern of results, responses would be more variable in random walk conditions compared to fully random. Therefore, reduction in variability for filled intervals does not closely reflect any of the fitted models. The exploration of a fixed strategy revealed that the two-state model provided the best fit for the filled durations.

Once the parameters were allowed to vary, the value of the transition parameter was required to be noticeably more than zero if the results were to show the best fit of an iterative model, in agreement with Glasauer and Shi (2019; 2021b). However, this transition parameter was close to zero. The final exploration of flexible strategies demonstrated a preference for a static model over an iterative one. Therefore, unlike recent research, an iterative model did not provide a good account of participant data, even when a filled reproduction method was used. Moreover, when the experimental conditions were designed to encourage an iterative strategy, a static model continued to provide a better fit.

2.4 General Discussion

The central aim of the current experiment was to gain an insight into how Bayesian priors update over time in the temporal domain. This aim produced the following questions: 1) If a single Bayesian strategy is used in all environments (fixed strategy), which provided the best fit, 2) Do participants alter their prior acquisition strategy according to certain conditions? To achieve this, Bayesian models with different assumptions about stimuli generation were fitted to duration reproduction data to see whether participants favoured a particular strategy. These Bayesian models included a static, iterative and two state model. Additionally, different stimulus environments were created in order to understand how they could lead to an improved fit for a different Bayesian model. The different stimulus environments consisted of a random walk vs fully random stimulus sequence and a multiple object vs a single object appearance. Three separate elements of the results will be discussed here. Firstly, the size of the compression magnitude and the amount of variability in responses allowed for the inference of a single Bayesian strategy by comparing to the model predictions. Secondly a discussion of the model fitting procedures which allowed for the exploration of flexible models. Finally, literature which has examined differences between filled and unfilled interval durations and their implications for the current findings will be discussed.

Compression magnitude and Bayesian model predictions.

In all conditions, there was under estimation of longer intervals and overestimation of shorter intervals demonstrating a central tendency bias that was likely to be driven by a Bayesian prior (Jazayeri and Shadlen 2010). This was shown in both Experiment 1 and Experiment 2.

For Experiment 1 (unfilled durations) specifically, the compression magnitude was smaller for random walk sequences compared to fully random sequences. This difference across conditions produced a significant main effect of randomisation. This influence of the stimulus features on the size of this bias towards the prior was an outcome that was in line with the predictions of an iterative model. Although when all Bayesian models were fitted to the data with fixed parameters across conditions, the two-state model demonstrated a physical appearance most like participants data. This was especially noticeable in the two-state model's ability to account for the size of the increase in compression magnitude for fully random conditions relative to random walk. In contrast, the size of this increase shown by the iterative model was much larger and therefore provided a less successful fit. Therefore, the difference in compression magnitude indicated that the two-state model provided the closest match to the data for unfilled durations. Although in this case the best fitting static model also predicted a small amount of reduction in compression for the random walk sequence relative to the fully random.

In Experiment 2, using filled durations, there was a main effect of object type. This suggested that participants weighted the prior more heavily (more bias) in intermittent (multiple) object conditions compared to permanent (single) object conditions. However, the exact reasons for this were unclear as the current research predicted the opposite direction of this effect, due to the potential suitability of an iterative model for tracking the properties of a single object over time. There is a need to reconsider this earlier assumption of the current research that objects viewed as a 'singular quantity' are more likely to be tracked and updated over time. This may not be true in relation to the temporal features of the stimuli. An alternative way to indicate the difference between single and multiple objects could be colour (another Gestalt grouping). In this way, grouping according to colour, instead of space could have provided a perception of single object (remains the same colour) or multiple objects (each disk appears as a different colour).

A main effect of object type contrasted with the main effect of randomisation that was found in Experiment 1, suggesting that different stimulus features might be influential in altering the amount of central tendency bias that was produced. Additionally, for the filled intervals, the size of the compression magnitude was similar across all conditions and this behaviour would have been predicted by a static strategy because for this model, differences in stimulus sequence are not relevant for prior formation. This is because there is no assumption concerning the stimulus on the previous trial like there is in an iterative model. Interestingly, an iterative model provided a poor match to the compression magnitude data in Experiment 1 and 2.

Variability of responses and model predictions.

Overall participants show much larger levels of variability than the models and this was likely to be because motor noise of responses were not fitted. The variability of participants responses also provided an insight into the best fitting model. This is because certain models predict a reduction in variability if their assumptions about stimuli generation were in line with how the stimuli were produced. For example, it would be expected that a model with iterative capabilities (iterative model or two state model) could reduce variability for the random walk conditions because it would have been well suited to the autocorrelated stimuli. This was not shown to be the case in Experiment 1. Instead, participant response variability for unfilled durations showed a significant interaction. This was followed up by a simple main effects analysis, which suggested that in intermittent object type conditions only, variability was reduced to a larger extent in fully random conditions compared to random walk conditions. This interaction indicated that a static model was more likely because this model predicts reduced variability for fully random conditions.

Experiment 2 (filled intervals) found a main effect of randomisation on the variability of participants responses, where the fully random conditions showed more variability. This finding would indicate an iterative strategy because it is minimising variability for the stimuli where the previous trial is highly predictive of the current trial. This is in contrast with Experiment 1 where the variability indicated that a static model was more likely.

Across the stimulus conditions, CV did not appear to vary in a pattern which followed the compression magnitude. Therefore, there was no clear relationship between bias (compression magnitude) and variability (CV). This means that a factor other than the sensory uncertainty was leading to the increase in prior weight. Or alternatively that participants were not acting in a Bayes optimal manner.

Flexible model strategy.

To investigate the possibility of a strategy that was flexible according to condition, the parameters of the two-state model were allowed to vary. However, even when provided with this freedom, there was little evidence to suggest that the parameters (v/r and q/r) were varying according to condition. The near zero values of the q/r parameter highlighted that, irrespective of changes to the stimulus condition, participants assumed a static prior distribution. This was true for both unfilled and filled intervals. The possibility of a flexible model strategy according to conditions was also explored by fitting either a static or an iterative model to each condition in isolation. In Experiment 1 and 2, the static model provided a better fit to participants at the group level compared to other models across all conditions. For unfilled durations, there was a small increase in preference for the iterative model for random walk conditions, relative to fully random conditions. However, the iterative model showed the worst fit of all three models in both experiments, which is line with the low values of the transition parameter when the two-state model parameters were allowed to vary across different conditions.

Differences between filled and empty intervals.

The fundamental difference between the two tasks in experiments 1 and 2 was how the participants were presented with and reproduced the stimulus durations. The ready set go (unfilled interval) task had a rhythmical aspect, as they are trying to perceive a demarcated interval between two points and recreate it exactly. It would be plausible that a rhythm becomes substantially easier to follow when the preceding trials are similar to the current trial than when they are dissimilar. This could explain why the compression was larger for the fully random conditions compared to the random walk conditions. The increased difficulty of the task could have manifested as an increase in weight being placed on the prior as a result of increased sensory uncertainty.

There may also be fundamental perceptual differences associated with using filled intervals (one continuous sensory image to be replicated), compared to using empty intervals (two markers, the duration between which must be estimated). The differences in perception of these two types of intervals have produced some mixed findings. Some experiments suggested that filled intervals were discriminated more successfully (Abel, 1972b, 1972a; Rammsayer & Lima, 1991) but in other experiments, the reverse was true (Simon Grondin, Meilleur-Wells, Ouellette, & Macar, 1998; Pfeuty, Ragot, & Pouthas, 2008). This mixed evidence from discrimination tasks makes it difficult to conclude which type of intervals were more difficult to discriminate, and as result, which intervals were perceived as containing more noise as a potential driver of this difficulty.

However, there appears to be some consensus surrounding a filled duration illusion whereby filled durations appeared longer to participants(Thomas & Brown, 1974). Longer durations are inherently noisier due to scalar expectancy (Rakitin et al., 1998). So, if filled durations appeared longer to participants, they are noisier than empty durations. Subsequently, from a Bayesian perspective they would have been perceived as containing more uncertainty (Jazayeri & Shadlen, 2010). If we were to assume this was the case in the current research, it would lead to a greater amount of central tendency bias for filled intervals compared to empty ones. However, in the filled duration experiment, the size of the central tendency bias for the filled durations was not consistently larger than the central tendency bias for unfilled durations. Furthermore, explaining the differences shown in the current research as arising due to increased uncertainty during filled interval reproduction also seems unlikely because participants reported on the difficulty of the empty interval task more frequently than they did during the filled interval task.

Conclusion

Overall, it was demonstrated that a static model, or a static like two state model were the most able to account for participants data. In the unfilled duration experiment, there was a suggestion of an iterative strategy shown by the empirical data containing reduced compression magnitude for the random walk stimuli compared to the fully random. This reduction in central tendency could have potentially been explained by the suitability of the iterative model to the random walk conditions and its ability to minimise over and under-estimation when stimuli were autocorrelated. However, the size of reduction in compression magnitude was more similar to a two-state model than the simple iterative model. This was also shown to be the best fitting fixed model when one potential strategy was explored for all conditions. Therefore, the fitting of the models to the empirical data in Experiment 1 showed that the iterative model demonstrated the worst fit. These findings were unexpected, and novel, given the merit of iterative models in recent research that was geared towards understanding how priors update over time. Glasauer and Shi (2019; 2021b) found that an iterative model was the best explanation of their data because participants showed reduced compression magnitude in response to random walk stimuli relative to fully random stimuli.

In the current research, the iterative model provides the least likely explanation for participant behaviour and whilst the reasons for this are not immediately clear it could be down to experimental differences (different lab rooms or instructions) or a theoretical requirement for unique strategies under certain circumstances that exists outside of experimental variables.

In an effort to understand the influence of experimental factors, and the discrepant findings between the current experiment and the findings of Glasauer and Shi (2019; 2021b) filled intervals were introduced in Experiment 2. However, in Experiment 2 the iterative model provided an even worse explanation of the data because there were similar levels of compression across the different stimulus order conditions (as predicted by a static model). Although the best fitting fixed model was shown to be the two-state model, closely followed by the static model, in agreement with Experiment 1. Moreover, there was not enough evidence to suggest that participants used flexible strategies in response to different stimulus features. Further research is needed to determine whether there are different stimulus features, experiment instructions or different participant groups, that could demonstrate the possibility of flexible strategy usage.

The theoretical implications of discrepant findings from different research groups are difficult to determine until further research is carried out. However, it could be that there is no single strategy for duration estimation exists across all tasks due to how temporal durations are indirectly accessed from the senses and contain lots of naturally occurring noise as a result. This is in contrast to spatial stimuli where direct sensory perception is possible via visual quantities (such as position and brightness). As a result, in spatial estimation a single strategy approach could be more likely, and this is explored in Chapter 4.

3.The impact of neurodevelopment on Bayesian magnitude estimation.

3.1 Introduction

The current chapter builds upon Chapter 2 by exploring the AQ trait data collected and the potential differences in biases, variability, and strategies for high AQ trait participants.

Introduction to perception in autism theories.

Empirical explorations of autistic perception have formed their foundations from non-social characteristics of including superior visuospatial skills relative to communication skills (Kumar, 2013), attention to detail (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001), and enhanced local processing (Mottron, Burack, Iarocci, Belleville, & Enns, 2003; Mottron, Dawson, Soulières, Hubert, & Burack, 2006). These perceptual differences have been explained by the Weak Central Coherence (WCC) theory, which suggests that autistic people show an advantage in processing the local details in contrast to typically developing individuals who prioritise global processing (global superiority; Happe, 1999; Happé & Frith, 2006).

Empirical support for the theoretical foundation of the WCC, local processing preference combined with a global deficit, is mixed (Simmons et al., 2009). Another theory, Enhanced Perceptual Functioning (EPF), proposed enhanced local processing in autism, with the lack of a global deficit (Mottron et al., 2003; Mottron, Dawson, & Soulières, 2009). However, not all autistic individuals possess superior visual processing. Furthermore, there is not only great variation in the research findings in perceptual characteristics, but the sensory processing underlying such perception can vary greatly from person to person. Depending on an autistic person's specific sensory profile they can experience both hyposensitivity and hypersensitivity. For instance, an autistic person may be hyposensitive to taste but hypersensitive to touch. This would lead them to seek out spicy or strong flavoured foods, but at the same time make an effort to avoid physical contact (Crane, Goddard, & Pring, 2009). The presence of these varied profiles begins to cast doubt on theories such as WCC and EPF because they do not provide the level of detail to explain the complexity present in autistic sensory experience. Taken together, a new framework for understanding the interconnecting mechanisms of sensory and perceptual characteristics of autistic perception is needed. Ideally, a framework that would be able to relate to other characteristics such as social communication differences and the need for routine (Van de Cruys et al., 2013).

The focus on local and global perceptual styles in WCC and EPF may provide an incomplete explanation of autistic experience. More comprehensive theories may be needed to account for mixed findings such as a lack of a global processing deficit during motion perception (Van der Hallen, Manning, Evers, & Wagemans, 2019) with some studies finding larger motion coherence thresholds in children with ASD than in typically developing children (Milne et al., 2002; Spencer et al., 2000) whereas many others showed comparable motion coherence thresholds (Del Viva et al., 2006; Jones et al., 2011; Milne et al., 2006; Price, Shiffrar, & Kerns, 2012). Plaisted and colleagues (1999) considered that the approach of certain tasks may have an impact on how autistic children display different global perception abilities to that of neurotypical children. More specifically, Navon figures, large letters comprised of smaller letters have been used to determine global and local processing preferences because they allow for the inspection of the attentional resources allotted to each type of processing respectively. On the basis of previous research using Navon figures, a divided attention (asking what the local and global letters were on each trial) task demonstrated a lack of global interference in autistic children (Bellville & Mottron, 1993) whilst Ozonoff and colleagues (1994) found there was a global interference effect using a selective attention (asking what the local or global features were on each trial) task (Ozonoff, Strayer, McMahon, & Filloux, 1994).

However, due to each of the above tasks using different groups of children Plaisted and colleagues (1999) aimed to compare the performance for same group of autistic and non-autistic children in both the selective and divided attention Navon tasks. The authors found that autistic children showed more errors for the global features in the in the divided attention task, but the non-autistic showed more errors in the local features (Plaisted et al., 1999). However, importantly the global advantage effect occurred for both non-autistic and autistic children during selective attention. This showed the global advantage effect to be something that was presented in both groups of children, when the task was set up to allow for the children to attend selectively to one feature level at a time. The global advantages displayed here alluded to the possibility that autistic perception may a show preference for local stimuli but only under certain conditions, hinting at the need for a mechanism which could explain the importance of these conditions.

Explanations of autistic behaviours may need to move beyond describing a detail oriented perceptual style in order to capture the relationship between perception and the complex array of characteristics displayed by autistic individuals such as restricted interests, need for routine and social communication differences. It is possible that these characteristics are driven by perceptual differences although weaknesses remain in all current accounts of autistic perceptual experience (Sapey-Triomphe, 2017). The following paragraphs of this introduction aim to explore some of the theories based on Bayesian decision theory and explain autistic characteristics as developing from weighting prior knowledge differently.

Highly detailed sensory perception is one possible outcome of weak priors in the Bayesian model (Pellicano & Burr, 2012). Potentially, preference for predictability because of a difference in the ability to weight and use priors could unite many seemingly unrelated facets of the autistic experience (Pellicano and Burr 2012). For example, the need for routine could therefore arise because priors are not being weighted to improve uncertainty in unpredictable environments. Additionally, social environments are unpredictable by nature therefore demonstrating a specific challenge for possessing weak priors in an uncertain environment. Pellicano and Burr (2012) were the first to develop a formal theoretical framework and this will be discussed in the following section.

Early Bayesian accounts of autistic experience: weak priors

Building on the WCC theory, Pellicano and Burr (2012) suggested that the global processing deficit contained within this theory could be formalised through a Bayesian observer model. This model describes how perceptual processes may be produced through the weighting sensory likelihoods and priors according to their relative uncertainty. In relation to autistic perception specifically, autistic perception may be experienced as ‘too real’ as the influence of prior knowledge, used by neurotypical individuals to improve predictive abilities, may be reduced. To produce this hyper realistic experience, Pellicano and Burr (2012) describe that weight given to the sensory likelihood information as relatively large and the weight given to prior information to contextualise sensory information as relatively small. The consequence of this would likely be sensory overload because prior knowledge is not being used to reduce noise that accompanies incoming sensory signals (Pellicano & Burr, 2012).

This Bayesian theory of autistic experience has been used to explain sensory and social differences (Pellicano & Burr, 2012; Van de Cruys et al., 2014). More specifically, weak (low precision) priors, could explain sensory overload due to the level of detail arising in perception free from the constraints of prior knowledge. Additionally, weak priors could produce social differences because in highly unpredictable interactions prior knowledge would usually provide context to subtle cues. Furthermore, increased weight given to prior knowledge would act to reduce the number of possible meanings down to the most likely meaning based on previous learning. Empirical support for weak priors was shown by the influence of prior knowledge in producing central tendency effects. The experiment carried out by Karaminis and colleagues (2016) involved temporal reproduction of a set of time intervals and the prior, of which the responses would be biased towards, was the mean of the stimulus distribution. Reduced central tendency bias in autistic children, relative to their sensory noise, demonstrated that they were not weighting the prior (mean of the stimulus distribution) as much as would be optimal in order to overcome the sensory uncertainty they displayed (Karaminis et al., 2016).

The weighted combination of likelihoods and priors has some crucial implications. For example, precise sensory likelihoods would appear in the same as weak priors in terms of outcome. Therefore, disentangling these processes becomes essential in understanding autistic perception as ‘too real’ because weak priors and precise sensory information could be equally plausible (Brock 2012; see 1.10, Figure 13). Recent autistic trait evidence supported a precise likelihood account. Karavelis and colleagues (2018) demonstrated that increased weighting of sensory likelihoods shown in high trait individuals as evidenced by weaker biases in combination with decreased variability in responses on low contrast trials in a visual motion perception task (Karvelis et al., 2018). If the weaker biases were not accompanied by decreased variability in responses, then this would indicate a weak priors account, and this is exemplified by the figure from Brock (2012). Whereas in Karvelis and colleagues (2018) reduced bias was found in conjunction with decreased noise in responses for participants with high AQ trait levels. Therefore, autistic traits may be related to the increased weighting of sensory likelihoods. This provided evidence for the possibility that increased precision of likelihoods could provide an alternative explanation to weak priors, even though weak priors and strong likelihoods would produce the same outcome from Bayesian integration (Brock, 2012).

Mottron and colleagues (2000) found a local bias in music processing in autistic individuals and Heaten and colleagues (2003) found that autistic people were better at remembering and labelling the pitch of musical notes. Taken together, showing how a detailed focused perceptual style with reduced bias may present in autistic individuals. However mixed support for the respective accounts of weak priors and precise likelihoods remains. Some tasks seem to show increased noise in autistic individuals' responses (Noel, Lakshminarasimhan, Park, & Angelaki, 2020; Karaminis et al., 2019), whereas others showed reduced noise and increased precision (Karvelis et al., 2018). A single explanation that could reconcile these accounts within a broader framework is needed. To achieve this, recent research has considered how priors develop over time. This could provide more mechanistic detail beyond purely analysing the weights on priors and likelihoods in isolation.

Highly inflexible precision of prediction error account (HIPPEA)

Some empirical evidence has, contrary to the weak priors account, displayed indications of strong and more precise priors from autistic and high trait individuals (Van de Cruys et al., 2017). The finding of Mismatch negativity responses (indication of surprise) occurring sooner in autistic individuals than neurotypical individuals supported the presence of strong (narrow) priors (Cléry et al., 2013; Gomot, Giard, Adrien, Barthelemy, & Bruneau, 2002). This is because for surprise to be shown, strong expectations must have been built up based on previous knowledge. Additionally, this was supported by Sapey Triomphe (2021) who presented autistic and non-autistic participants with two different ranges of disk sized stimuli. Autistic individuals continued to show narrow prior standard deviations (strong priors) irrespective of whether the stimuli were from the narrow or wide range and as a result demonstrating a lack of flexibly adjusting the width of their priors to suit the stimulus context (Laurie Anne Sapey-Triomphe, Timmermans, et al., 2021). Therefore, testing at with different tasks could lead to seemingly strong or weak priors depending on the exact conditions for prior acquisition. This means that accounts of autistic perception must go beyond a sole focus on the size of the weight placed on priors and likelihoods and instead emphasise how priors might be sensitive to changes in sensory inputs over time. To achieve this, more recent research has explored how previous sensory information may be accumulated as priors form and update according to contextual factors.

Several factors may be involved in how and when how sensory information is used to update priors. According to a predictive coding (a type of Bayesian) framework, prediction error is used to determine how much each sensory signal should be allowed to cause a prior update. In this framework, prediction error replaces sensory likelihoods and the prediction itself replaces the prior (Van de Cruys et al., 2017). When prediction error has high precision, then this indicates a new statistical regularity to be learned and the prior should be updated. However, when sensory information is noisy (more of a one-off change to the statistical regularities), then precision of prediction error is reduced so that is does not lead to a prior update. This is believed to be the process that occurs within neurotypical individuals as it allows for the flexible adjustment of prediction error precision according to sensory signal changes. Therefore, making updates to the prior will allow them to change their model to keep in line with how the world changes. However, it was suggested by Van de Cruys and colleagues (2017) that autistic individuals might have fixed inflexible high precision of prediction errors. This would lead to overfitting and learning according to highly specific sensory input. This makes generalisation-based learning difficult because a single specific situation is unlikely to repeat itself.

Hierarchical theories

Another potential prior updating mechanism is hierarchical learning. The hierarchy here referred to how prior weighting occurs on multiple levels, with the lower level concerning prior weighting where the source of the uncertainty is the sensory signal itself. Whereas the higher-level only referred to beliefs about the source of the uncertainty that is manipulated by the experiment. This means that, at the higher level, sensory information is treated differently according to the stability of the certainty (volatility). Behrens and colleagues (2007) used a learning task to demonstrate that in environments where the reward probability was unstable, the previous information from more recent trials is used to inform learning decisions. However, when the reward probability is stable then, participants use the previous experimental outcomes across many trials to guide their learning (Behrens et al., 2007). The operation of learning rate implies that when the world is stable, learning rate needs to reduce because there is nothing new to be learned (Dayan, Kakade & Montague, 2000). There are occasions where an unstable reward environment can signal a new statistical regularity. Once the new regularity is learned, this provides a more optimal prior in this fluctuating environment.

This mechanism may explain how neurotypical individuals adjust their learning in response to environmental fluctuations (Manning et al., 2017; Palmer et al., 2017) however, there is some evidence to suggest that autistic adults overestimate volatility (Lawson et al., 2017) and this may lead to a consistently increased learning rate. This overestimation of volatility was determined from a perceptual measure that uses pupil dilatation to indicate surprise. Although the same level of overestimation in perceiving volatility may not necessarily be visible in these individuals at the level of reward learning. The learning task used involved showing the children two treasure chests on each trial with a certain probability that the ‘correct’ chest would be chosen. The initial phase was a familiarisation, that allowed the children to learn and 80:20 ratio of reward probability (for 20 trials). Then during the experimental phase, the stable condition provided a consistent 75:25 ratio. Finally, in the unstable condition the probability alternated between 20:80 and 80:20 every 20 trials. Manning and colleagues (2016) found no difference in learning rate between autistic children and typically developing children. Although there was also the possibility that Manning and colleagues (2016) learning task was too easy and so created a floor effect, disguising any differences between autistic and typically developing children. Additionally, the study that found an overestimation of volatility was carried out with autistic adults. Therefore, the possibility remains that despite overestimation of volatility not being visible in children, this process could develop through adulthood as a way of coping with the persistent unpredictability of life.

Recent theories imply Kalman model.

Both theories (HIPPEA and hierarchical) share some core features which means that they predict a prior acquisition strategy akin to the operation of a Kalman filter. The HIPPEA theory explained that autistic individuals may form priors in such a way that involved high inflexibility of their prediction error precision and as a result learn many different fluctuations in the environment, even if it reflected noise rather than a consistent, learnable statistical regularity. As a result, this theory would predict a Kalman filter style model for prior acquisition because changes to the stimulus are continually learned on every trial, even if they are very small or highly inconsistent with what came before.

Hierarchical theories similarly assume that volatility is overestimated by autistic individuals (Lawson et al., 2017). Overestimation of volatility is reflected in higher learning rates than necessary when the environment is relatively stable (Palmer et al., 2017). The outcome of high learning rates is that participants would be ready to learn a change to the prior value more quickly than if they had a slower learning rate. This is not directly measured in the current research however, if one of the two models were to facilitate a higher learning rate (and short term learning) it would be explained by the Kalman filter model because its prior acquisition relies on looking to the value of the previous stimulus. This would contrast with a static model which learns a mean of all the stimuli presented and therefore would provide a basis for long term prior learning. Taken together, both theories explain mechanisms which prioritise more recent information (but via slightly different processes), and therefore predict a different prior updating process to the static model.

Evidence against recent theories implies static model.

In contrast, recent research suggests that global information, such as the mean of the stimuli presented so far is given more weight by autistic individuals than to local (previous trial) information. Lieder and colleagues (2019) aimed to understand how previous trial information was used by autistic individuals. More specifically, whether the global (mean of all stimuli presented) or local (the previous trial stimulus) information is more able to predict the current response magnitude. They used a generalised linear model (GLM) to do this. The model consisted of bias by long term (average of all stimuli) and bias by short term (average of N-1 and N-2). It was found that for autistic people, the bias by the long term information was equivalent to typically developing individuals (TD). However, the bias by short term was reduced in autistics compared to TD individuals. Implying that autistic individuals favour using the global information over local because the global information better predicted their response.

This finding contrasts with the explanation of autistic experience provided by hierarchical and HIPPEA theories which would predict a preference for more recent information. This is because the increased learning rate in autistic individuals hypothesised by the hierarchical theory implies that recent information is closely attended to and used to adjust prior updating. Alternatively, a model that would more closely align with Lieder and colleagues (2019)'s findings would be a static model (where the M and SD are updated) which places less emphasis on the value of more recent trials.

Therefore, the current experiment explored Bayesian prior updating strategies. Bayesian computational models were fitted to see whether prior updating is different for individuals with a higher level of autistic traits compared to those with a lower level of autistic traits. The current research focuses on fitting models according to different generative assumptions (how participants thought stimuli were generated) to find out which assumption is more likely held by autistic individuals: either 1) that the previous trial is most informative or, 2) that the summary statistics (M and SD) of all previous trials is most informative.

Current research

Using data from previous temporal reproduction experiments, any biases of magnitude estimation were correlated with AQ questionnaire data. The aim of this correlation was to explore whether the amount of compression bias varied according to autistic traits. The HIPPEA and hierarchical accounts described so far could both lead to more of a Kalman filter type model. This is because they produce the perception that learning needs to be updated more frequently. However, it is also possible that, according to the findings of Lieder and colleagues (2019) a static Bayesian model would provide this best fit to the data.

The current experiment used data from previous chapters including compression magnitude (central tendency bias measure), coefficient of variation (variability measure) and Akaike model weights (best fit measure) from a temporal reproduction task. These measures will be used to explore individual differences in temporal perception using high and low autistic trait groups. From these theories and their previous exploration, we can make some predictions about AQ traits as a measure of individual differences:

As a result of the theoretical background discussed up until this point, a negative correlation between compression magnitude and AQ would indicate weak priors. As was the case in previous chapters, a reduced compression in responses to random walk conditions relative to fully random sequences could indicate the use of an iterative model. Although, as before models were fitted to the trial by trial data in order to better understand how successfully each model described the data. A further prediction concerned the correlation between the variability of the responses (CV) and the AQ score. If a negative correlation between AQ and compression magnitude was accompanied by a negative correlation between CV and AQ, this would suggest that the seemingly weak priors were actually driven by high sensory precision. Akaike weights were used to assess the model’s likelihood as they were in in previous chapters. A negative correlation between the likelihood of the static model and AQ would suggest previous trial as an autistic strategy and therefore support HIPPEA (Van de Cruys 2013) and volatility (Manning et al., 2017; Palmer et al., 2017) based theories. However, a positive correlation between the likelihood of the model for the static model and AQ would imply a static model as an autistic strategy, in line with Lieder and colleagues (2019).

3.2 Methods

Participants

In the unfilled experiment, there were 26 (13 female) participants from the ages of 20 to 43. One participant had to be excluded due to outliers. In the filled experiment there were 51 participants with an age range from 18 to 53. All but 4 participants were students at the University of Nottingham. Normal or corrected to normal vision was a requirement of all participants. From the original experiments carried out in Chapter 2, 16 participants did not provide AQ scores. The data from these participants was not used in this analysis, leaving a total of 59 participants.

Data for the values of compression magnitude, CV and Akaike weights was pooled across participants from Experiment 1 and 2 of Chapter 2. To create a mean average compression magnitude value for each participant, the values of (unsigned) compression magnitude calculated at each of the 13 durations were averaged. This same process was carried out on the CVs and Akaike weights to produce an average CV and Akaike weight for each participant. The values displayed random walk and fully random sequences were kept separate but averaging across single and multiple object types.

Procedure

The basic procedure involved temporal reproduction of the stimuli as described in Chapter 2. A single run of each condition lasted for 5 minutes meaning that the time taken to complete the entire task was 40 minutes.

There were two separate experiments (between subjects) with two different types of stimuli; filled interval stimuli (see Figure 18 in Chapter 2) and unfilled interval stimuli (see Figure 25 in Chapter 2). For the filled interval stimuli participants were asked to press down the button for the length of time the stimuli were on the screen for.

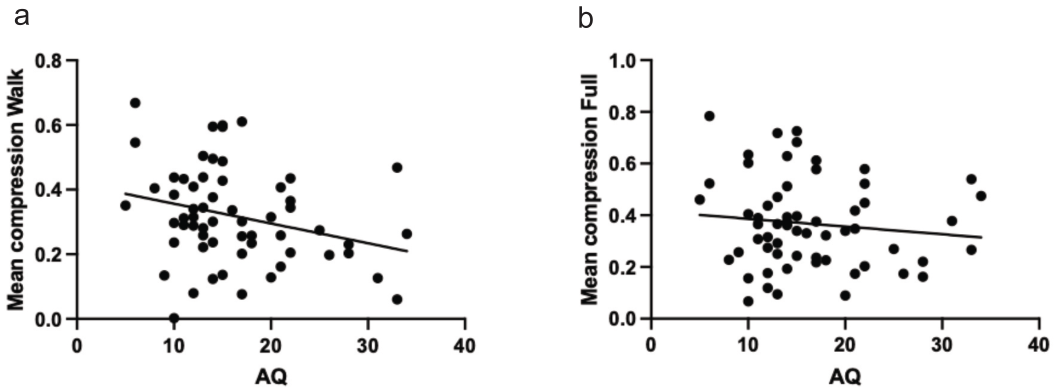
Whereas in the unfilled interval experiment participants were asked to press the button when they felt that the amount of time passed between their button press and the second stimulus was equivalent to the amount of time between the first and second stimulus. In both temporal reproduction experiments, the temporal sequence of the duration stimuli varied such that random walk sequences were used for two conditions and fully random sequences were used for the other two conditions. Performance metrics were also averaged across permanent and intermittent object conditions. In addition to the change in temporal sequence across conditions, Additionally, the AQ questionnaire was given to participants after completing temporal reproduction task. This led to a total experiment time of 45-50 minutes.

3.3 Results

The mean compression for each participant was calculated by averaging the compression magnitude (see Chapter 2 for how this was calculated) in the two random walk conditions (permanent walk and intermittent walk). The mean CV value for individual participants was correlated with AQ scores in Figure 27A. The same process was used to calculate the mean compression across the fully random conditions (permanent full and intermittent full) and correlate these values with the AQ scores (see Figure 27B).

The mean compression in the random walk conditions shows a negative correlation so those with a higher AQ score showed lower compression and this correlation was significant, r (58) = -0.2775, p=0.032). In the fully random conditions, the correlation was weak, negative and was not statistically significant, r (58) = -0.1171 p= 0.373.

Figure 27. Mean compression correlated with AQ scores.



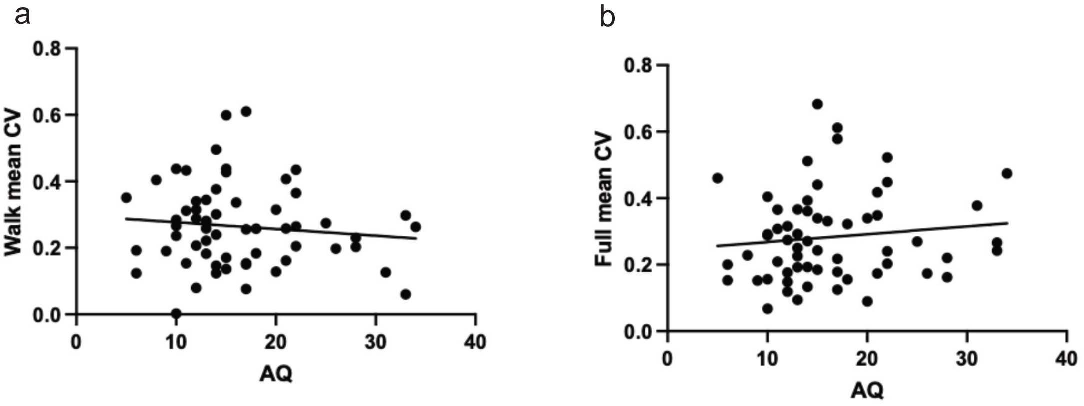
Note.

The 1-b compression magnitudes correlated with the AQ scores and fitted with a linear regression to display the strength and direction of the correlation. The mean compression magnitudes plotted here were averaged across a) all random walk conditions (two object types ; permanent and intermittent and two stimulus types; filled and unfilled) b) all fully random conditions.

The mean CV for each participant was calculated by averaging the coefficient of variation (see Chapter 2 for how this was calculated) in the two random walk conditions (permanent walk and intermittent walk). This mean CV participant value for random walk conditions was correlated with participant AQ scores in Figure 28A. The same process was used to calculate the mean CV across the two fully random conditions (permanent full and intermittent full) and correlate these values with the AQ scores (see Figure 28B).

The correlation between mean CV and AQ in the random walk conditions was negative and non-significant, r(58) = -0.1098, p = 0.404). Whereas the corelation between the mean CV and the AQ in the fully random conditions was positive. This was also non-significant r(58)=0.1174, p=0.372 .

Figure 28. Mean coefficient of variances correlated with AQ scores.



Note.

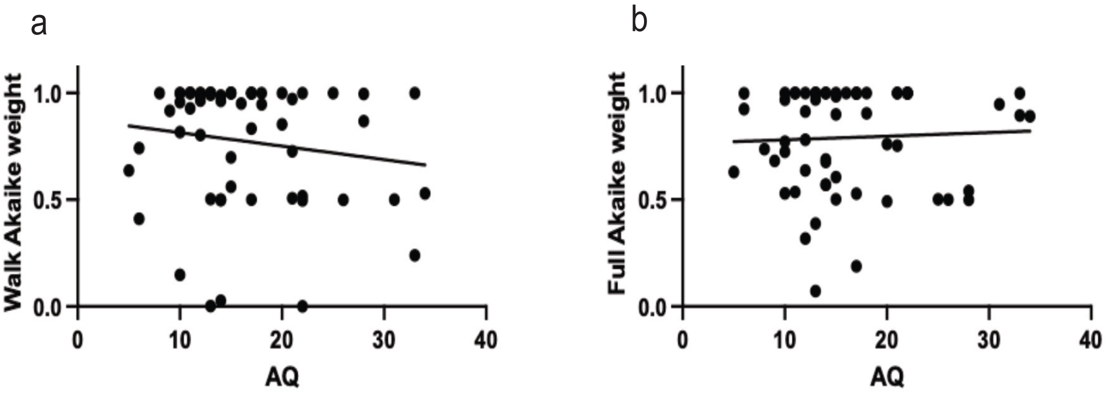
Coefficient of variances (CV's) were correlated with the AQ scores and fitted with a linear regression to display the strength and direction of the correlation. The CV's plotted here were averaged across a) all random walk conditions (two object types ; permanent and intermittent and two stimulus types; filled and unfilled) b) all fully random conditions.

Figure 30 shows the mean Akaike weights for all participants. The individual Akaike weights, which determined the likelihood of using each model, were correlated with AQ scores. An Akaike weight of 1 means it is likely they are using the static model as opposed to an iterative model (a value of 0). See Section 2.2.3 to for details of how these weights describing the most likely model were calculated for each participant.

The mean Akaike weights were produced by averaging across all the random walk conditions (filled and unfilled stimuli and their presentation as intermittent and permanent object types; Figure 29A) and all the fully random conditions (Figure 29B). A large amount of participant Akaike weights had a value of 1. There are some participants at the lower end of the Akaike weight but not enough for there to be a clear linear relationship between AQ and the extent to which a static model may be used. The preference towards a static model is clear across the majority of participants (as seen in Chapter 2).

The correlation between the mean Akaike weight for the static model was negative and non-significant in the random walk conditions r(58) = -0.1474 p= 0.261. In the fully random conditions, the mean Akaike weight in relation to the fit for the same model was positive and non-significant r(58) = 0.04804, p=0.715.

Figure 29. Mean Akaike weights correlated with AQ scores.



Note.

The mean Akaike weights are correlated with the AQ scores and fitted with a linear regression to display the strength and direction of the correlation.

The Akaike weights plotted here were averaged across a) all random walk conditions (two object types ; permanent and intermittent and two stimulus types; filled and unfilled) b) all fully random conditions.

3.4 Discussion

Mean Compression

The current research showed a decrease in compression magnitude (central tendency) as AQ score increased (both in negative direction in random walk and fully random). This meant that high trait individuals showed relatively reduced compression compared to low trait individuals in both conditions. Although, the negative correlation between compression magnitude and AQ was, larger in the random walk compared to the fully random conditions. The negative correlation in the random walk conditions was also significant whereas it did not reach significance in the fully random conditions.

The significant negative correlation for high trait individuals appears to be in line with the suggestion made by Pellicano and Burr (2012) of reduced prior weighting (weak priors) in autistic individuals. Although more recent research has detailed that this account does not explain how priors are learned differently in autism (Sapey-Triomphe et al., 2021; Van de Cruys et al., 2013). In terms of detailing the mechanisms at work, a reduction in compression could be indicative of many possibilities that lead the prior to be weighted less. For example, it has been demonstrated that priors were weighted less under conditions where the precision of sensory likelihoods are high. Drummers with low sensory variability were able to show veridical perception with reduced weighting of the prior, as evidenced by reduced central tendency effects (Circhinni et al., 2012). More specifically, they also found that as variability, as measured by CV, increased so too did central tendency bias (prior weighting). Therefore, demonstrating the possibility that the increased bias towards the prior occurs when responses contain more noise.

Reduced compression in random walk conditions relative to fully random stimuli is an outcome that would be predicted by a simple iterative model (Glasauer & Shi 2021). Due to the negative correlation between compression magnitude and AQ in the random walk conditions, there is the possibility that high trait individuals may be using an iterative model. Although as will be discussed later, this was shown to be less plausible after the likelihood of the iterative model was assessed using Akaike weights.

Theisinger et al., (2022) also explored the size of central tendency effects in relation to stimulus sequences (random walk and fully random) and autistic traits. Similar to the results shown in this chapter, they did not find convincing support for a correlation between AQ scores and central tendency effects. However, the diagnosed autistic group showed a significant difference in the size of the central tendency effect compared to typically developing individuals (a reduction in bias for autistic individuals). Furthermore, when viewing random walk sequences autistic participants showed the lowest bias to the mean across all groups and sequences (Theisinger, 2022).

Mean CV

Participant variability, in the form of the CV was correlated with AQ traits in order to investigate the possible implications for the precision of sensory likelihoods in autistic experience. If the reduced weighting of the prior occurred as a result of increased precision of likelihoods, then a decrease in participant variability measures may be seen. Participant variability may reflect the precision of the sensory likelihoods to an extent (despite containing other forms of noise such as motor noise). However, if there was a real relationship between likelihoods and AQ, then it would be expected that overall participant variability would reflect this. If there is not a decrease in sensory variability, then weak priors may provide a better explanation as described by Pellicano and Burr (2012).

In the case of the current research, an increased precision of likelihoods would require a negative correlation between mean CV and AQ scores to be shown. However, this correlation was only shown to be negative in the walk condition and was weak (and not significant) in both conditions. Therefore, sensory estimations were not more precise in high AQ trait individuals and the reduction in the weight assigned to the prior (compression magnitude) cannot be ascribed to increased precision of likelihoods.

Mean Akaike weights

A further possibility of reduced central tendency bias towards a mean, specific to the random walk conditions could be the use of an iterative model. This model, when used for both types of stimulus sequences predicts a reduction in compression for random walk sequence stimuli (Glasauer & Shi, 2021b).The operation of this model is successful in accounting for the amount of random noise from small changes in the random walk stimuli, therefore there is less need to over and underestimate in line with the central tendency bias in comparison to randomised stimuli.

In order to explore the ability of static and iterative models to describe the data from the current research Akaike weights were calculated. The weights themselves vary between 0 and 1 for representing the likelihood of the iterative and static models respectively. The sum of their weights was equal to 1 and as a result, correlations of AQ scores could be carried out with either Akaike weight scores as the weight for the static model would be the inverse of the iterative and vice versa. In this case, AQ scores were correlated with the Akaike weight for the static model. The correlation of the AQ scores with the Akaike weights (averaged for walk and full conditions) demonstrated no clear linear relationship between the model used and the AQ score of the participant. Most participants, irrespective of AQ score appeared to show a static model and there were many weights equal to 1. This means that the generative assumption from the used by participants, via the static model, was that the stimuli were drawn from a distribution.

AQ cut off and other limitations.

One possible explanation for a lack of significant correlations could be that we did not have many high trait individuals. Scoring 32 or more out of a possible 50 was argued to be an informative cut off (Baron-Cohen, Wheelwright, et al., 2001). This was later verified by Woodrow Smith et al., (2005) with an argument for 26 with GP referral but 32 in the general population with no other evidence. In the current research there were only 3 individuals scoring over the cut off. The AQ as an instrument is also not able to capture the full experience of having the condition especially in terms of the precise sensory characteristics. It would likely need to be used in combination with another questionnaire in order to understand this dimension of participants behaviour (Robertson & Simmons, 2013). It does however measure related cognitive processes such as displaying high attention to detail and poor attention switching and these constructs have the potential to capture a detail focused cognitive style as described by Weak Central Coherence. Moreover, the AQ has been praised in its ability to accurately predict diagnoses from 80% of people over the 32 cut off, whilst only 2% of general population show AQ scores greater than the cut off (Baron Cohen 2001).

Iterative versus static models

As was highlighted in the introduction there are a few recent schools of thought on how prior knowledge may be incorporated in autistic perception. Firstly, Lieder and colleagues (2019) found that autistic individuals were more likely to use the average of all previous trials to inform their current response. This suggested a similar mechanism to the prior updating of the static model in the current research. The findings of the current research were such that for nearly all participants, the static model was the most likely model.

In contrast, two other Bayesian autism theories have argued that autistic experience may involve learning differently in response to more recent sensory evidence. One such account, which focuses on fixed high precision of prediction error (HIPPEA), explains that if precision of prediction error is too high, then updates to learning are made very frequently and precisely. The product of learning in relation to specific instances is that priors become difficult to generalise and they are sensitive to noisy fluctuations. Another theory which seemed to focus on how recent learning was different in autistic individuals is the volatility overestimation theory, which explains that autistic individuals may show an increased learning rate, relative to the amount of actual volatility in the environment. This is defined as an increased learning rate because it would be advantageous to decrease the learning rate as an environment becomes more stable and predictable. Both theories would predict that learning or prior updating occurs closely to recent information, similar to the action of the iterative model in the current research. However, these theories were not supported by the current research, because the iterative model would not a likely model for high trait individuals specifically as its suitability was already low across the full sample.

Conclusion

Mean compression appeared to show a relationship to AQ traits such that there was a weak but significant correlation in the random walk conditions. Although in the fully randomised sequence conditions, the weak negative correlation was not significant. This provided support for priors becoming weaker as the number of autistic traits increased only in the case of random walk sequence stimuli and therefore did not provide support for a weak priors account. Due to there being no visible signs of decreased variability as measured by the CV, the increased precision of likelihoods was not supported.

The mean Akaike weights showed that the static model as the most likely and this was not related to AQ score. Lieder (2016) found that autistic individuals were more likely than non-autistic individuals to use the entire distribution (akin to the static model when making estimates about frequency of tones. In the current experiment, there was not a clear difference in the use of static vs iterative models according to autistic traits. Moreover, a static model which learns the distribution of the stimuli, was favoured by most participants. Therefore, an iterative Bayesian model was unlikely and, as a result, theories which could rely on this type of mechanism, such as HIPPEA and overestimation of volatility were not supported in the current research. However, due to the low numbers of high AQ trait individuals (AQ >32) the preference for a static model is unlikely to be representative of autistic perception.

4. Bayesian modelling of spatial localisation task behaviour: flexible prior updating in response to different stimulus sequences.

4.1 Introduction

The findings of Chapter 2 revealed the preference from a static model in duration estimation. This may have been for several reasons but most likely because temporal durations may be perceived as discrete and uncorrelated magnitudes. The current chapter aimed to explore the possibility that autocorrelated stimuli could encourage iterative strategies for spatial quantities. Another stimulus difference that was present between the experiments in Chapter 2 and Glasauer and Shi (2019; 2021b) was a variable step random walk (drawing a random step size value each time instead of using a fixed step value). This adjustment to the stimulus sequence was brought forwards to the current chapter. Bayesian integration has been supported through a range of spatial domain phenomena such as shape from shading (Adams, Graf, & Ernst, 2004), visual motion perception (Weiss et al., 2002), path integration (Petzschner, 2012), spatial localisation (Bejjanki, Knill, & Aslin, 2016; Berniker, Voss, & Kording, 2010; Kiryakova, Aston, Beierholm, & Nardini, 2020) and motor control (Berniker & Kording, 2009). According to Bayesian theories, magnitude estimation biases, which at first glance appear to be errors, are in fact systematic ways in which prior knowledge is combined with noisy sensory estimates to improve their reliability (Jazayeri & Shadlen, 2010).

As demonstrated in Chapter 2 and throughout the wider literature, central tendency effects have provided a quantifiable and robust index of the weight placed on the prior (Jazayeri & Shadlen, 2010; Petzschner & Glasauer, 2011; Petzschner et al., 2015). These effects have also been used more recently to demonstrate the rapid formation of priors over the course of an experiment (Roach et al., 2017). A Bayesian framework has been used to interpret central tendency bias when estimating quantities such as the length of a line (Ashourian & Loewenstein, 2011) and the duration of time intervals (Jazayeri & Shadlen, 2010). In the area of duration estimation specifically, participants estimates became biased towards the mean of previously presented intervals and participants' showed the best fit to a Bayesian (Bayes least squares) model (Jazayeri

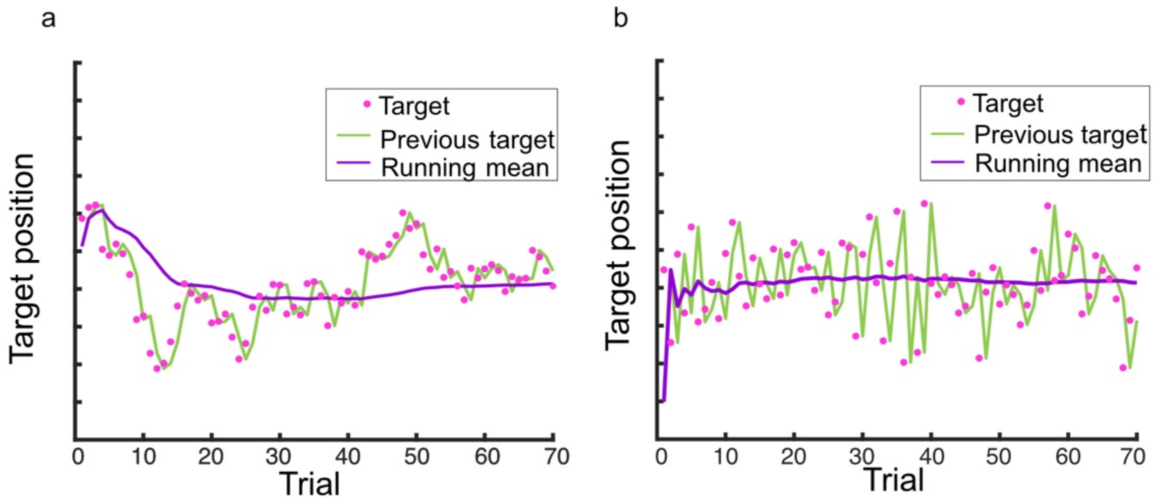
& Shadlen, 2010). Despite these promising findings, research still needs to explore 1) how these strategies operate at the trial level and 2) if these strategies can be used in a flexible way according to different learning environments. Initial explorations of these ideas began by considering the possibility that the central tendency bias as an outcome could arise as a result of the inherent uncertainty of random presentation. This meant considering the extent to which over and under estimation could be caused by the experimental protocol of randomly presented stimuli. Reduced central tendency bias may occur if the stimuli were presented in a way that changed gradually over time rather than sporadically. In addition, naturally occurring quantities in real life environments such as speed or sound levels tend to change gradually. This contrasts with the sporadic shifts between different discrete values produced by fully random presentation (Glasuer & Shi 2019; Glasauer & Shi 2021b).

Glasauer and Shi (2019;2021b) demonstrated a reduced central tendency bias in responses when stimuli were presented via a random walk sequence compared to fully random presentation. The random walk sequence involved the selection of a duration at random for the first trial only and then thereafter an equal chance of a one-step increase or decrease (with some Gaussian noise) in the size of the duration. Reduced bias with random walk sequences was likely found because participants were using a Kalman filter strategy, a strategy well suited to gradually changing stimuli because prior updating occurs according to the most recent changes in the stimuli. As a result, the model predicted a decrease in central tendency bias, and the participants estimation behaviour also displayed this decrease, but it was not as large as the model predicted. Overall, they concluded a fixed Kalman filter was the best fitting model but were, at this point, unable to explain this difference in compression between the model and the data. Considering that different levels of randomness were present in their two stimulus orders, it also seems plausible that participants could have switched strategies. Although this was not the case for the duration reproduction experiments displayed in Chapter 2 where the strategy used by most participants was a static Bayesian strategy regardless of the stimulus sequence used to present the stimuli.

There is the possibility that different strategies may be suited to different types of stimuli, due to the unique advantages that a given prior acquisition strategy could provide for certain stimulus features. A particularly important characteristic of prior acquisition and updating might be the order (randomness or lack thereof) in which stimulus usually appear in the learning environment. Psychology as a discipline often favours random presentation however, this is not always how real life environments produce learnable statistical regularities. Fully random stimuli, if drawn from the distribution and, if their values do not influence one another are described as independent and identically distributed. A real-life example of this might be how long a weekly meeting could potentially last. There might be some factors that influence fluctuations in the length of time but crucially the time of one meeting does not directly dictate the length of any other single instance.

In contrast, there are certain environments where each instance of the event is related to the next (autocorrelated). An example of this would be the background noise at a restaurant. Often when one person starts talking more loudly, another person will have to speak over that noise level to be heard (the same applies for decreases in noise level) meaning that each volume level can be predictive of the next. This real world example functions similarly to the statistical regularity provided by a random walk sequence because there is a small, but random, incremental increase or decrease in magnitude. More crucially this means that the size of the previous stimulus magnitude, is predictive of the current stimulus magnitude. This is the same type of stimulus sequence used by Glasauer and Shi (2019; 2021b).

Figure 30. Prior updating strategies: previous target and running mean for random walk and fully random sequences.



Note.
The shift in target positions from Gaussian distribution (M = 0, SD = Walk SD) plotted as a function of the trial number. The filled dots show the target positions for a) the random walk sequence b) the fully random sequence. The green line shows the previous target prior, and the purple line shows the running mean prior in both figures.

To better contextualise these prior acquisition strategies and the benefits they provide for certain stimulus sequences, see Figure 30. For the random walk type of sequence, the small changes in the stimulus provide an environment where the current target position can be easily predicted from the previous target position (autocorrelated). Therefore, the prior updating strategy which accumulates prior knowledge through learning the stimulus target location on the previous trial would be most successful (Figure 30a). However, for a fully random sequence, then it is advantageous to use the average of these large fluctuations to provide a summary of the data because here, the previous target position is highly different to the current target (Figure 30b). Taken together**, if** different real-life environments inherently contain different levels of randomness, it would be plausible that different prior updating strategies would be better suited to each of these environments. Additionally, the results of Chapter 2 from this thesis raise the possibility that there could be certain qualities of temporal durations which make them less likely to be suited to a Kalman filter strategy. To explore these possibilities, the current experiment aimed to understand whether, in response to different stimulus presentation order, which model provided the best explanation of participants behaviour during a spatial localisation task.

Spatial localisation tasks have presented virtual cue and target coins in a statistically related way to see whether humans use Bayesian perception to estimate their location. This is known as the 'Catch the Coin' task (Tassinari, Hudson, & Landy, 2006). The task involves the presentation of a cue coin, which disappears after a few seconds. This is shortly followed by an on screen net which needs to be dragged by the participant to the location where they think the hidden target might be. Finally, the position of the hidden coin is revealed, and the feedback of the hidden position provides to some extent, prior knowledge of the target coin position for future estimates. The full prior learning strategy could involve learning the mean of the distribution of the target coin positions (running mean) or by using the target coin location on the previous trial (Kalman like filter). The cue coin positions, however, formed the likelihood because they provided current sensory knowledge at the time of responding. This is because the cue coin positions were drawn from a Gaussian distribution (a normal distribution where the mean was equal to the target coin on that trial). According to Bayesian interpretations, when the standard deviation (variability) of the cue coin was large, the weight given to the cue coin reduced and the weight given to the prior increased (Bejjanki et al., 2016; Berniker et al., 2010).

Initial research using the 'Catch the Coin' task (Berniker et al., 2010) demonstrated that participants responded in line with Bayesian weighting of the prior and likelihood according to their reliability. Bayesian weighting was shown because as the prior variance decreased, the participants perception of the target location (net response based on where they believed the target to be) moved towards the prior mean. Whereas when the likelihood variance decreased, participants beliefs about the target location (net placement) moved closer to the cue coin, meaning that they were weighting the likelihood (cue) more heavily. Additionally, irrespective of whether prior variance was constant, or it switched half-way through the blocks (from 0.2 to 0.05), participants learned the prior mean very quickly. They also learned the prior variance, but over a slower timescale (Berniker et al., 2010).

Other research in support of these claims demonstrated that the respective means and standard deviations of multiple prior distributions can be learned simultaneously (Bejjanki et al., 2016). More recently, Kiryakova and colleagues (2020) explored Bayesian transfer in order to separate Bayesian perception from potential heuristic strategies, where transfer would not be possible. Bayesian transfer, in the case of this experiment, involved participants weighting the prior more heavily as they moved onto high variance likelihoods (relative to low and medium variance likelihoods they had experienced from the training). Some transfer was found because participants weighted the prior to a greater extent very early on during these post training test trials, which were not viewed during training. This indicated that a Bayesian process is likely to be in operation. Overall spatial localisation research seems to suggest prior and likelihood distributions were learned fairly successfully during this task, however there is a consistent finding of overweighting the likelihood and underweighting the prior relative to optimal Bayesian learning (Bejjanki et al., 2016; Kiryakova et al., 2020).

In the current experiment, data were collected across a range of cue and target standard deviations in order to gain insight into how the central tendency biases varied according to different sources of variability (and possible Bayesian integration) in this task. It was predicted in line with Berniker and colleagues (2010) that as the cue standard deviation increased, the bias towards the mean of the Gaussian distribution that the target positions were drawn from would also increase. This would indicate the presence of Bayesian integration.

Secondly, based on Glasauer and Shi (2019;2021b)'s findings it will be predicted that, if the participants are using a fixed Kalman filter like (previous target) strategy for both conditions, they will show reduced compression in the random walk condition relative to the fully random conditions. This difference would indicate the use of a previous target model because this type of prior updating is very well suited to the gradually changing stimuli. Alternatively, that participants could use a distribution estimation strategy (running mean), where the same amount of compression would be visible in both conditions because the value of the mean and standard deviation would be the same irrespective of order.

It is also possible that participants may be flexible in their prior updating and use a Kalman filter like (previous target) model to improve performance in random walk conditions, and then switch to a distribution estimation strategy (running mean) when it would be appropriate for learning about the stimulus properties during fully random conditions. To address this possibility, the current research aimed to use computational modelling to gain insight into the strategies that may have been elicited by different stimulus orders. So, in addition to collecting empirical data, three (two Bayesian models and one non-Bayesian) models were used to determine which strategy was being used when participants were presented with different stimulus orders. The Bayesian models included one that assumed the Kalman filter like learning (previous target) and one that assumed distribution estimation (running mean).

4.2 Methods

Observers

20 (10 males) observers took part in the research. Of those who disclosed their age, there was an age range of 21- 44. Participants did not report any differences in their vision so it was assumed that they were wearing appropriate eyewear. One observer was excluded due to some technical difficulties with running the online experiment on their personal computer.

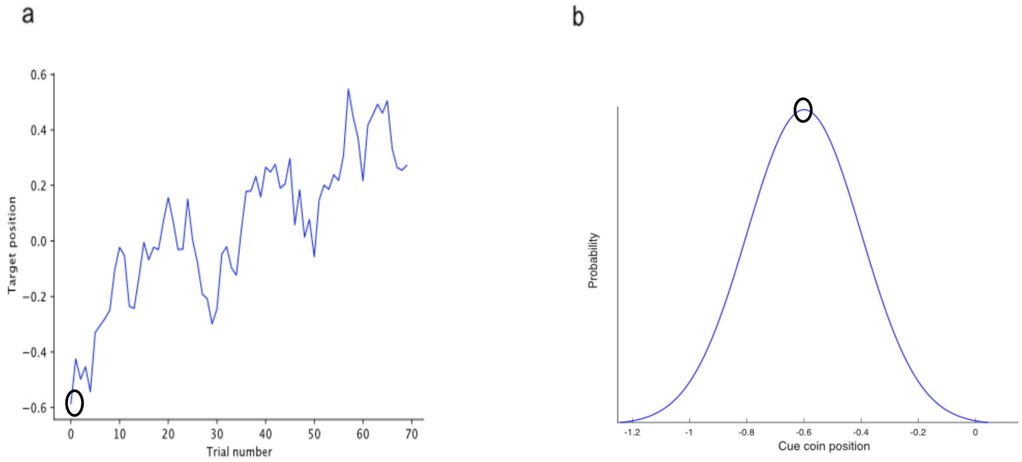
Stimuli details

The stimuli were presented on the participants own computer, using Psychopy Pavlovia due to the online nature of the experiment. However, the screen dimensions were standardised to a 4:3 displaying monitor to produce a width of -0.67 to 0.67 screen units and a height of -0.5 to 0.5 screen units. Psychopy software was used for experiment creation and the online platform it provided, Pavlovia, used GitLab (Peirce et al., 2019).The cue coin and the target coin were both circular in shape, coloured yellow and red respectively to distinguish them. They had a radius of 0.05 screen units. The white rectangle that was used to make the net was 0.07 screen units wide, and full screen height.

Procedure

During the experiment, the participants saw a cue and a target coin appear on the screen. They appeared in succession and were described to be 'being thrown from behind them' and landing in a virtual lake stimulus on the screen. In actuality, they were appearing and disappearing on screen, but previous spatial Bayesian integration research has used the narrative of coins being thrown to encourage participants to view the coins as landing in a predictable manner. A fixation cross appeared on screen at the start of the experiment for 0.5 seconds, then the cue coin followed and remained on screen for 0.3 seconds. The cue coin was a noisy indicator of the true target location because it was randomly drawn from a Gaussian distribution. This cue Gaussian had a mean equal to the target location on that trial (see Figure 31a and b) and a standard deviation (cue SD) that dictated how helpful the cue was on each trial. The generation of each target position involved randomly selecting a value for the first position followed by subsequent random walk step changes to this value for the remainder of the sequence. This random walk step was drawn from a Gaussian with a SD (equal to the walk SD for that condition) to determine the size of the variable step. A small cue standard deviation would lead to a more helpful cue due to there being reduced variability of the cue coin's position.

Figure 31. The generation of the cue coin positions from the target coin positions.



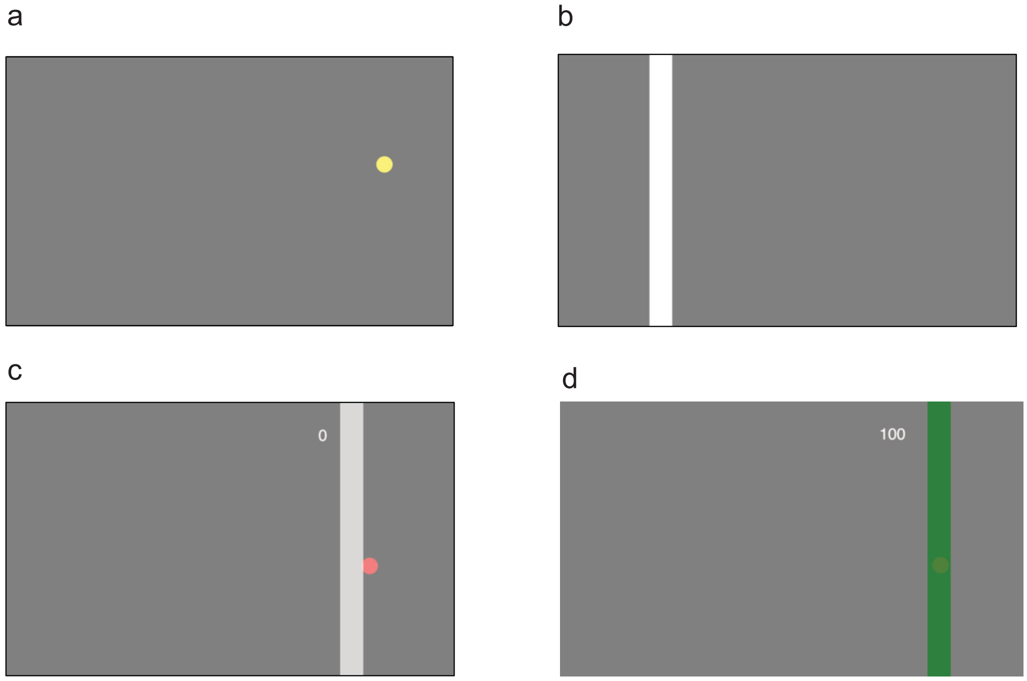
Note.

a) The target position path plotted as a function of trial number.

b) The cue coin position was drawn from a Gaussian with a SD equal to cue SD. The mean of the Gaussian the cue coin was drawn from was centred on the current target coin position. In the example shown here, the first target coin position is -0.6 in Figure 31a and as a result the cue coin Gaussian is centred on -0.6. The standard deviation of the cue Gaussian in this example is 0.2 which was chosen arbitrarily for the demonstrative purposes of this figure.

The participants were then asked to predict where the second target coin would ‘land’ based on where they saw the first one ‘land’ on the screen (see Figure 32a). The participants were presented with a net and then had unlimited time to use the mouse to place the net where they thought the target coin was located (see Figure 32b). After they confirmed their estimate by pressing space bar, the target coin was revealed and appeared on the screen for 0.75 seconds. If the participant moved the net’s centre to be less than 0.05 screen units away from the target coin’s centre, then the net turned green to indicate that they had “caught” the coin (see Figure 32d). This sequence was repeated for 70 trials to form one condition.

Figure 32. Sequence of events for a single spatial localisation trial



Note.

a) The yellow cue appears first b) The white rectangular net has mouse contingent horizontal movement c) If the target coin is not caught, the net turns translucent grey d) If the target coin is caught, the net turns green, and 100 points is added to the score for each coin that was caught.

Conditions

Fully random versus random walk

The order of the trials was fully random in two of the conditions and a variation on random presentation, random walk, was used to present the target coin in the other two conditions. Producing the target positions involved selecting a starting position at random and then using variable step increments to produce a random walk. The step size was determined by drawing a number from a Gaussian (a normal distribution curve with a mean of 0 and a SD equal to the predetermined value of the random walk SD). The walk SD therefore added unique variability to the size of each of the walk steps. The random walk of the target coin positions was constrained between -0.67 and 0.67 screen units, to avoid targets appearing off the screen. To create the fully random conditions, the corresponding random walk sequence was generated and then shuffled so that the order of the trials was the only difference between the two sequences.

Walk SD varying and cue SD varying.

To understand the different levels of compression according to the variability of the cue and the random walk sequence, either the cue and the walk standard deviation (cue SD and walk SD respectively) took on a range of values whilst the other of the two values was held constant at 0.05. This means that for two of the conditions the cue SD cycled through five values (0.01, 0.05, 0.1, 0.15 and 0.2) whilst the walk SD remained at 0.05. The sequences generated with these cue SD values were presented in blocks so 70 trials would be presented with a cue SD of 0.01 and this process repeated for the next 4 cue SDs. For the other two conditions, the walk SD varied according to these values and the cue SD remained at 0.05 (presented in blocks as described for the cue SD varying). Alongside the walk and full sequence order variations, this created 4 sessions; cue SD varying and random walk presentation, cue SD varying and fully random presentation, walk SD varying and random walk presentation and walk SD varying and random walk presentation. Each of the 4 sessions (350 trials) lasted for 30 minutes, leading to a total experiment time of 2 hours (1400 trials). To reduce the impact of practice effect, 10 participants were presented with the 4 sessions in one order (cue SD varying and random walk, cue SD varying and fully random, walk SD varying and random walk, walk SD varying and fully random) and the other 10 participants experienced the tasks in the reverse order.

4.3 Data processing and model details

The participants' mouse (contingent with a net stimulus) position was used to record the estimates of the target coin position. The amount of bias in participants estimates was measured using the difference between their estimate of the target position (net position) and the predicted net position (mean of target coin positions). To gain this insight into participants biases towards the target mean (the prior, equal to 0 in this case), the cue coin position was plotted against their net response position and these scatter plots were fitted with a linear regression.

The word compression was used to describe how the linear regression fit line appeared flattened (a gradient closer to 0 than it was to 1). This flattening, according to Bayesian theories, occurs due to how participants' responses become biased towards the prior, where the prior is the mean of the target coin positions. Under conditions where b (the gradient) was low, compression effects were large. Repurposing the value of compression magnitude as 1-b made the compression more intuitive because then high values of 1-b suggested compression effects were large (Roach et al., 2017). Compression magnitudes were calculated for each individual participant, in each condition. These values were then averaged and plotted against the standard deviation of cue and the standard deviation of random walk step respectively to see how bias changed as the variability of these stimulus properties was altered. Random walk and fully random conditions for each cue and walk SD were plotted on the same figure in order to provide a comparison for the different amount of compression shown for the two stimulus orders.

The visual features of the linear regression that is, how flattened the fit line appears, provided an indicator of how much the participants were biased towards the mean of the target locations. If the net was placed directly in line with the cue coin (likelihood), then the gradient, b, of the linear regression line would be 1. If, hypothetically, the maximum amount of compressive bias towards the mean of the target positions was shown, the gradient would have been equal to 0. The following simple models were devised in order to see which of the strategies would provide a best fit to participants responses. The no prior (cue coin only) model is described first and then two different Bayesian models which produced a combined weighting of the cue and target are described.

No prior (cue only)

The simplest strategy would involve participants using only the cue on the current trial to aid their estimate. This non-Bayesian model has no prior knowledge of the target and therefore, this model was set up by assigning a weight of one to the cue. The net locations produced by this model were represented by each net location being equal to the cue location on the current trial.

Dynamic prior: Running mean.

Another possibility is that participants are learning the mean of the target location distribution and are updating them according to this calculated mean, up to and including the most recent trial. This model involves placing the net according to a weighted average of all the target positions seen so far $w\overline{T_{1:trial}}$ and the cue on the current trial C_{trial}. As a result, it updates the mean in a dynamic way to incorporate the value of the most recent trial. It can be summarised by the following equation:

$$N_{trial} = w\overline{T_{1:trial}} + (1 - w)C_{trial}$$

Where w represents the weights, and the size of this weight determines the amount of compression. A weight of 0 would be no compression and a weight of 1 would be the net placed on the mean of all the target positions up until that point on each trial.

Dynamic prior : Previous target

Alternatively, it could be that participants rely on the most recent known location of the target coin. This model involves updating the prior according to the target position on the most recent trial. The net position then becomes a weighted sum of the target position on the previous trial $wT_{trial-1}$ and the cue coin on the current trial C_{trial}. It can be summarised by the following equation:

$$N_{trial} = wT_{trial-1} + (1 - w)C_{trial}$$

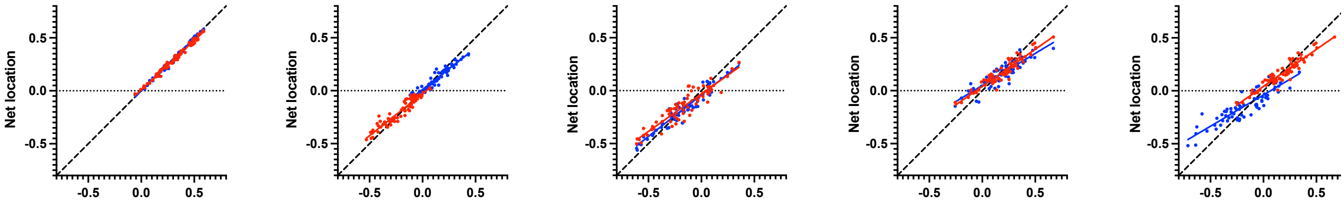
The fit of each model is likely to depend on the conditions presented to participants, given that the running mean and the previous target models may be suited to different stimulus environments. To investigate this possibility further, the models were fitted to the participants responses and the Akaike weights were used to assess the fit, where a larger Akaike weight indicated a higher likelihood of the model being used.

4.4 Results

Cue SD: single participant

Figure 33 shows example scatter plots of the net location and the cue coin location as the cue SD increased for a single participant. This provided a situation where if the net was placed on the cue every time, it would give the gradient a value of 1 (the dotted line). As the standard deviation (SD) of the distribution that the cue coin positions could be drawn from became increasingly noisy, the walk SD remained fixed at a value of 0.05. This participant showed an overall increase in compression magnitude, as the cue SD increased from 0.01 to 0.2 in random walk and fully random conditions. This systematic increase was shown more consistently in random walk conditions.

Figure 33. Empirical data showing the cue coin location plotted as a function of net (response) location as cue SD varied.

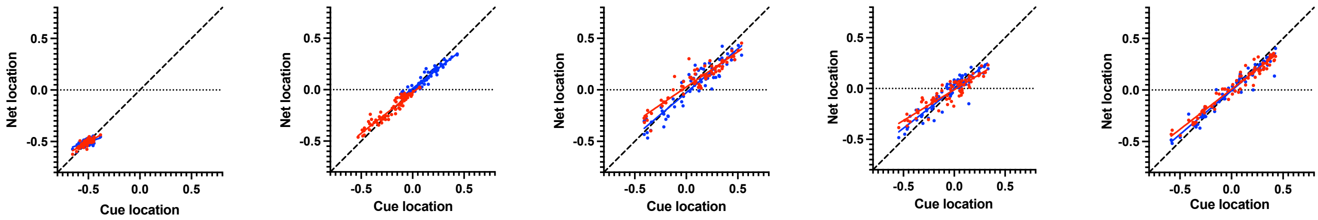


Note. These data were fitted with linear regressions to show the amount of bias displayed by participant 4. The cue SD varied from 0.01 to 0.2. Walk SD was fixed at 0.05. Blue dots and blue fit lines indicate random walk conditions and red dots, and red fit lines indicate the fully random condition. The dotted line represents the case where participants place the net on the cue coin on every trial.

Walk SD: single participant.

In the walk SD varying conditions, the cue SD was a fixed value (0.05). Figure 35 shows example scatter plots (fitted linear regressions) for each walk SD condition from a single participant. Again, the net location was plotted against the cue location. At the lowest possible walk SD (0.01) there was an extremely large amount of compressive bias shown by the linear regression fit appearing almost horizontal. Increasing the walk SD (from 0.01 to 0.2) led to a clear decrease in compression. For this single participant, there was slightly greater compression magnitude shown in the fully random conditions compared to the random walk conditions for all walk SD's excluding walk SD = 0.01.

Figure 34. Empirical data showing the cue coin location plotted as a function of net (response) location as walk SD varied.

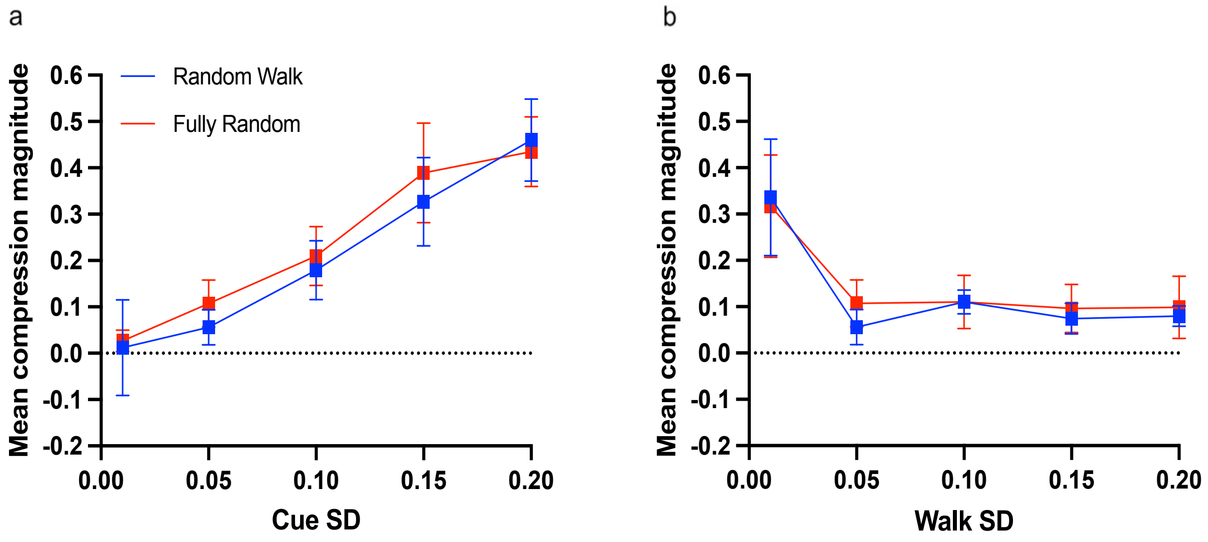


Note. These data were fitted with linear regressions to show the amount of bias displayed by participant 4. Walk SD varied from 0.01 to 0.2. Cue SD was fixed at 0.05. Blue dots and blue fit lines indicate random walk conditions and red dots, and red fit lines indicate the fully random condition. The dotted line represents the case where participants place the net on the cue coin on every trial.

Mean compression magnitude

Group averaged compression magnitude was subsequently plotted as a function of cue SD (see Figure 35). In line with the individual participant data, as the cue SD increased, the mean compression magnitude also increased. Mean compression magnitude was smaller in random walk conditions compared to fully random conditions, excluding trials where cue SD = 0.2 where the reverse was true. A repeated measures ANOVA was carried out to assess whether the varying the cue SD and changing the stimulus order demonstrated any effect on the compression magnitude. The first ANOVA was a 5 (cue SD = 0.01, 0.05, 0.1, 0.15, 0.2) x 2 (order = walk, full) ANOVA and produced a significant main effect of cue SD on compression magnitude, F(1,18)=74.799, p<0.001, eta squared=0.631. Although, there was no main effect of stimulus order F(1,18)= 1.026, p = 0.325, eta squared = 0.004 and no interaction F(1,18)=0.760, p=0.555, eta squared =0.006

Figure 35. Mean compression magnitudes for the spatial localisation task



Note.

a) The mean compression magnitudes of subjects as the cue SD varied from 0.01 to 0.2. Walk SD was fixed at 0.05.

b) The mean compression magnitudes of subjects as the walk SD varied from 0.01 to 0.2. Cue SD was fixed at 0.05.

Once again consistent with the individual participant data, as the walk SD increased, the mean compression magnitude decreased. Mean compression magnitude showed a reduction in compression for all random walk conditions compared to fully random conditions, excluding the condition where walk SD was equal to 0.01; see Figure 35b). However, this did not manifest as a main effect of condition in a repeated measures ANOVA. This second ANOVA was another 5 (walk SD = same values as cue SD) x 2 (order = full, walk). Similarly, a main effect of walk SD on compression magnitude was found F (1,18)=20.492, p<0.001, eta squared=0.397 but no main effect of stimulus order F(1,18)=0.986, p=0.334 eta squared = 0.005 and no interaction F(1,18)=0.696, p=0.597 eta squared=0.006. Post hoc comparisons using the Bonferroni correction revealed differences between walk SD set to 0.01 and all other walk SDs (p<0.001).

4.5 Computational Modelling

The simple model simulated the experimental procedure to reveal which Bayesian weighting strategy (as described in the earlier model details section) would provide the best fit for the participants response pattern. The Bayesian models (running mean and previous target) were fitted to the individual trial data using the function 'fmincon' in MATLAB. The cue only model was used for the purposes of understanding which model may be more likely but could not be fitted to the data because it had no prior weight parameter. This model fitting process involved allowing the free parameter, prior weight, to vary. The empirical prior weight was produced when the difference (sum of squared error) between the data and model was minimised. The empirical weights show under weighting of the prior compared to optimal weights.

The optimal prior weight was the weight that used Bayes rule to provide a weighted sum of the prior and likelihood according to their relative uncertainties (Figures 36 and 37). In this case, the uncertainties of the likelihood and the prior are provided by the values of the standard deviation of the cue and the standard deviation of the prior respectively.

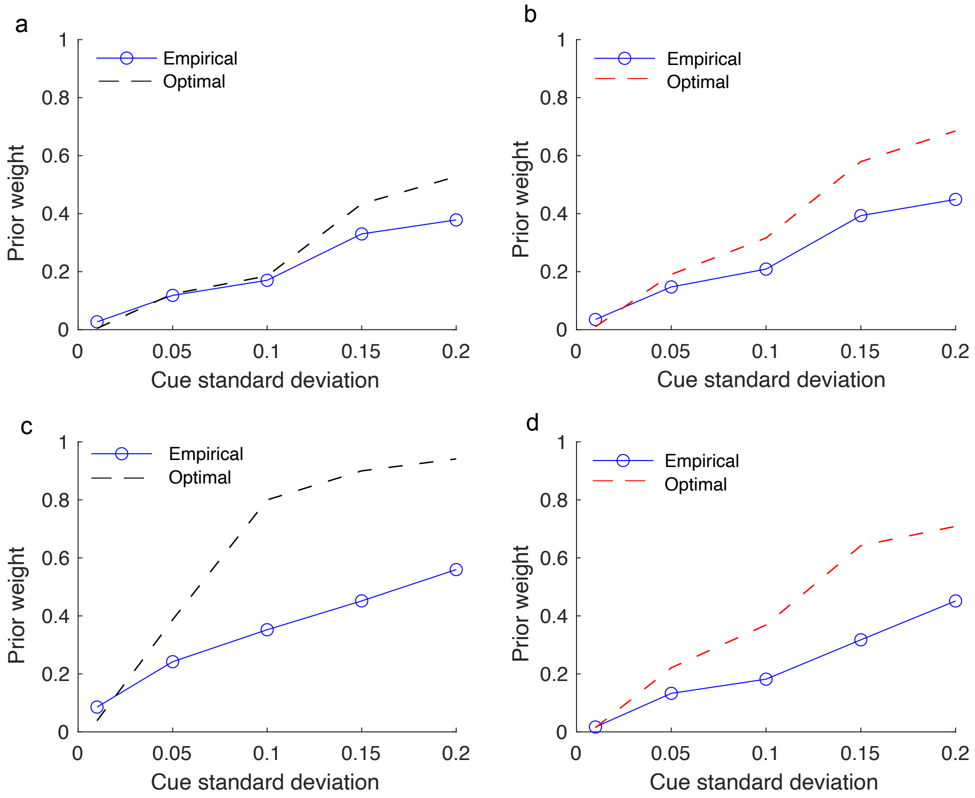
For the previous target model, the optimal prior weight was equal to:

$$\frac{{}^{1}/_{\sigma_p^2}}{{}^{1}/_{\sigma_p^2} + {}^{1}/_{\sigma_l^2}}$$

Where σ_p was equal to the prior standard deviation (walk SD) and σ_l was equal to the likelihood standard deviation (cue SD). When the random walk sequence was shuffled (fully random conditions) the prior standard deviation was in fact equal to the standard deviation of the target positions presented in that block.

In the running mean model the main difference was that the prior was learned according to the mean of the stimulus block (mean of the 70 trials). Additionally, a running estimate of the standard deviation was maintained and updated by calculating the prior SD on each trial for random walk trials. For fully random trials, the prior standard deviation was again equal to the SD of all the target positions.

Figure 36. Empirical data and optimal weights plotted as a function of the cue SD.

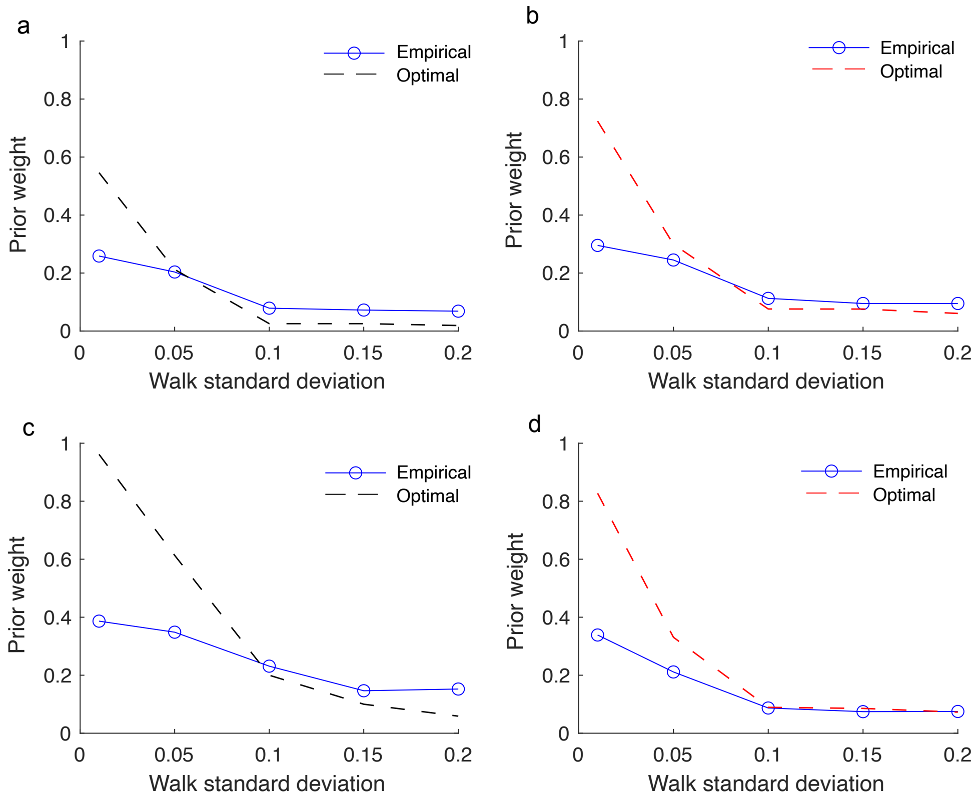


Note.

For a) the previous target model and full trials, b) the running mean model and full trials, c) the previous target model and walk trials, d) the running mean model and walk trials.

Figures 38 and 39 show the mean compression produced from the best fitting prior weights in the case of fitting both sequences simultaneously (Figure 38) versus fitting to each stimulus sequence (random walk and fully random) separately but plotted on the same figure (Figure 39). The modelled net responses for each participant were produced using the best fitting individual prior weights mentioned in the previous paragraph. The amount of compression shown by the model fits was calculated and the average of the compression magnitude for the fitted data was plotted as a function of the SD that was varying for that condition (cue SD or walk SD).

Figure 37. Empirical data and optimal weights as the walk SD increased.

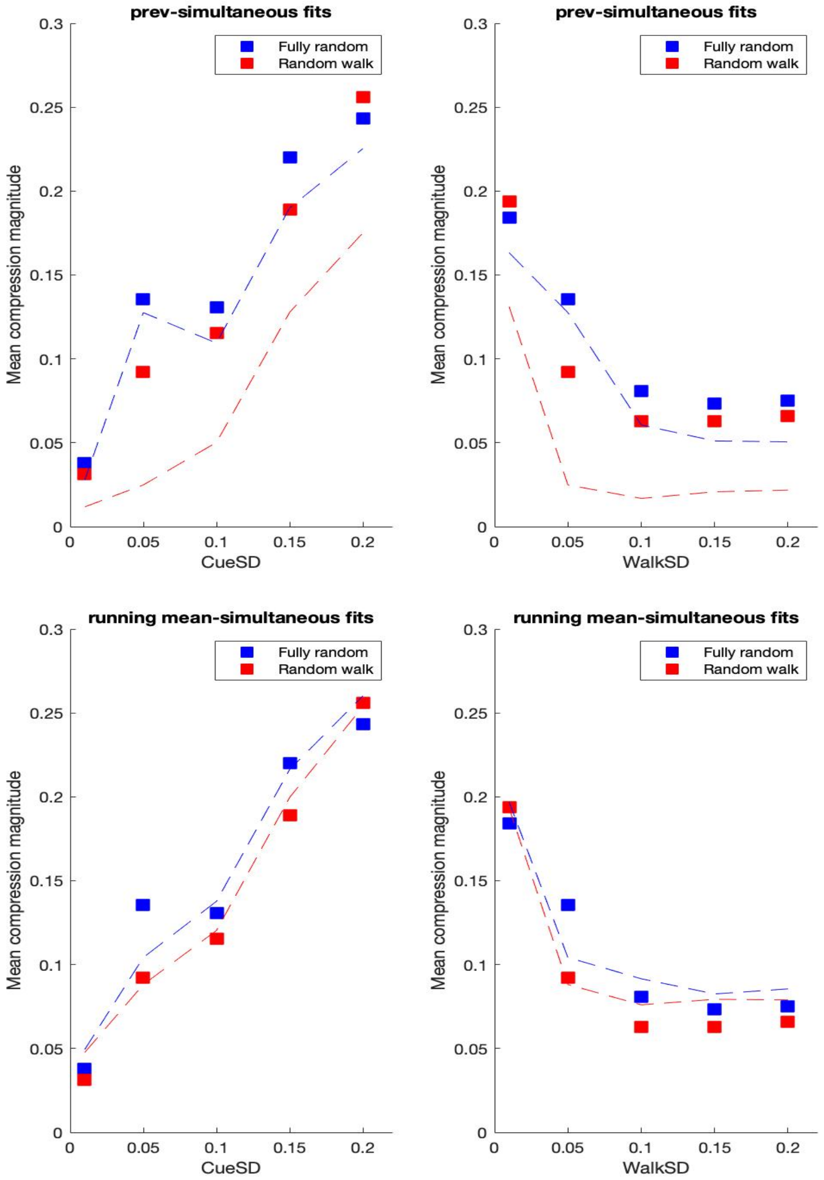


Note. For a) the previous target model full trials, b) the running mean model and full trials, c) the previous target model and walk trials, d) the running mean model and walk trials.

Initially, both types of stimulus sequence were fitted simultaneously (see Figure 38) in order to demonstrate the best fitting strategy across all conditions. The first modelling process of simultaneous fitting assumed that a single fixed model was used by participants across both conditions. In Figure 38 it was evident that the no prior (cue only) model shows a very poor fit of the data and so a Bayesian strategy was more likely than a cue only strategy. As for both fitted Bayesian models, there were some visible similarities to the data including how the prior weighting increased as cue SD increased. However, in contrast to the data, the previous target model predicted large reductions in compression for the random walk sequences relative to the fully random sequences. The data showed no main effect of randomisation so this may have been expected. Further detail concerning the quantitative assessment of these fits will be discussed in the following paragraph. In addition to these simultaneous fits, the prior weight was then allowed to vary separately for each condition. This meant that different best fitting models (and prior weights) could be shown for each condition in turn (see Figure 39 for the separate fits for each condition plotted on the same figure). Again, the no prior model shows a poor fit to the data. Although the separate model fits provide a much better description of the data than the simultaneous fits shown in Figure 38.

The mean compression magnitude calculated from the model fits provided a coarse lens for comparison to the empirical data because it was unable to show how the prior was acquired over time and how models may show varied success in fitting this prior updating process. To provide a more detailed analysis of trial by trial fits and to assess the quality of those fits the Akaike's information criterion (AIC) and the resulting Akaike weight were calculated (see Section 2.2.3 for Akaike weight calculations). This criterion considers the quality of the model according to the residual sum of squares but also penalises it according to the number of free parameters it has. It is expected that more free parameters could allow for a better fit but over fitting should be avoided because it reduces the model's generalisability. The AIC values revealed which model was the best fit to participants data in each condition, with smaller values indicating a better fit. Additionally, an Akaike weight was calculated from the AICs to provide a more meaningful quantity for comparison that is, the likelihood of that model being used. The weights act probabilistically meaning that the larger the weight, the more likely it is that the model was used and the sum of all the weights will be equal to 1.

Figure 38.The mean compression for the simultaneous model fitting process for previous trial and running mean models.



prev-simultaneous fits

prev-simultaneous fits

running mean-simultaneous fits

running mean-simultaneous fits

Note.

The model was fitted to each participant on every trial. Fitted net responses were produced by prior weights (empirical) which minimised the SSe. The compression from the fitted net values was calculated and averaged as it was for Figure 35.

The mean compression is plotted here as a function of the cue and walk SDs for the respective conditions in the same manner as the data in Figure 35.

After the 3 models (running mean, previous target and no prior) were fitted to each participant individually then the residual sum of squares error, AIC's and respective Akaike weights were averaged across participants (see Table 1). As before, the first aim was to determine if participants were using one fixed model, which provided the best fit. To this end, a single model was fitted to both the random walk and fully random trials simultaneously. The Akaike weight of 0.5879 demonstrated that the previous target model provided the best fit in this case.

Table 1. The fit assessment measures calculated as the mean across participants for a single fixed model.

Model	Residual sum or squares	Akaike information criterion (AIC)	Akaike weight	Number of free parameters
No prior	10.39548291	7003.243456	1.30496E-36	0
Running mean	6.942883923	7548.260691	0.412064143	9
Previous target	6.642280074	7602.439907	0.587935857	9

As described in previous paragraphs, the subsequent model fitting process involved allowing the free model parameter (prior weight) to vary according to the condition (separate model fitting). It may be possible that different models could provide the best fit for the random walk and fully random conditions respectively and so it was necessary to consider what might happen if the model was fitted to only the random walk conditions or, only the fully random conditions. When only the random walk conditions were fitted the highest Akaike weight was seen in the previous target model (Akaike weight = 0.8328, see Table 2) and conversely when only the fully random conditions were fitted, the running mean model demonstrated the highest Akaike weight (Akaike weight =0.7323, see Table 3). Therefore, different strategies were a better fit to the different stimulus orders.

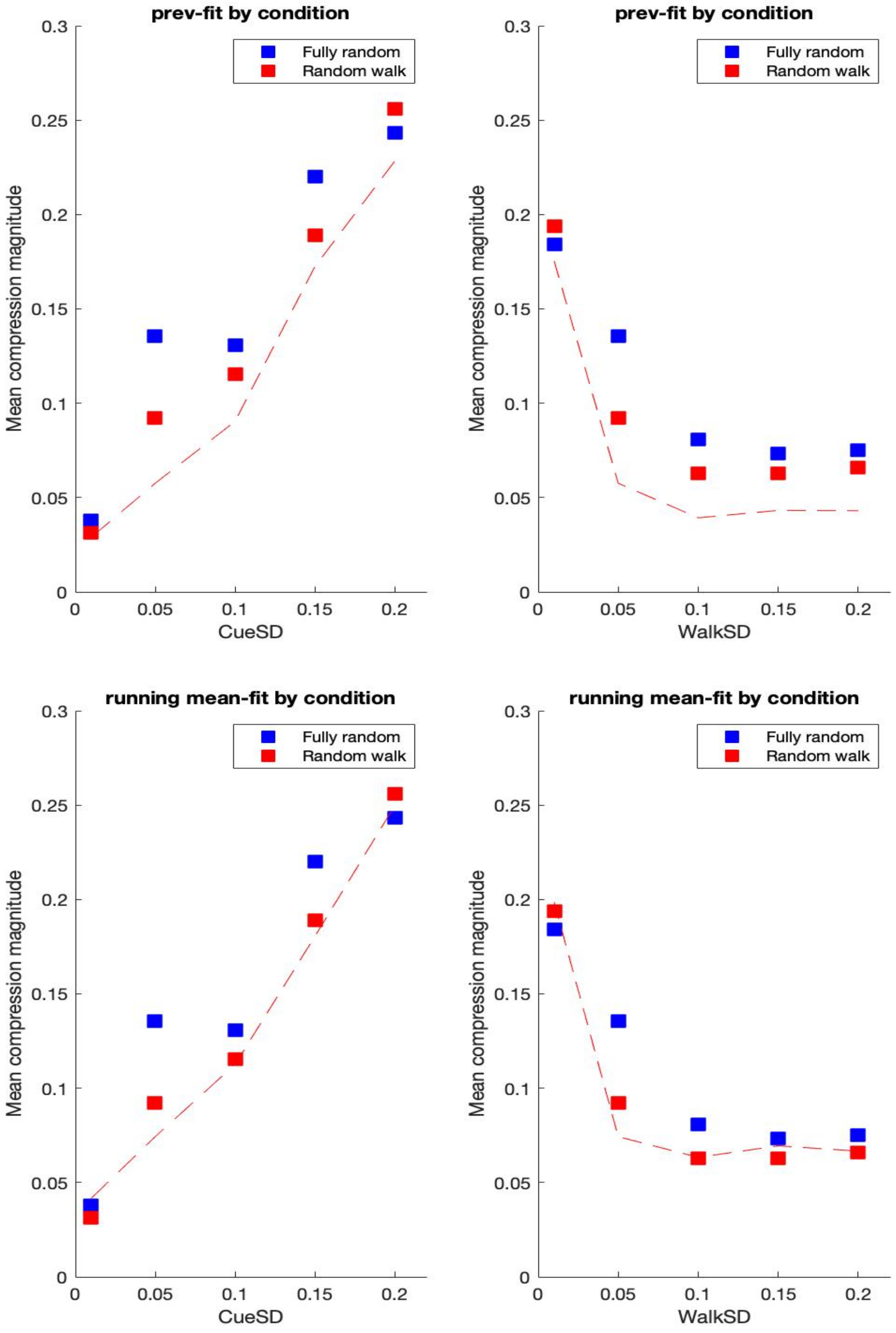
Table 2. The fit assessment measures calculated as the mean across participants for parameter fits of the random walk conditions.

Model	Residual sum of squares	AIC	Akaike weight	Number of free parameters
No prior	6.133699196	3520.861668	0.040953126	0
Running mean	4.057059825	3807.121115	0.126240627	9
Previous target	3.611071917	3993.514012	0.832806247	9

Table 3. The fit assessment measures calculated as the mean across participants for parameter fits of the fully random conditions.

Model	Residual sum of squares	AIC	Akaike weight	Number of free parameters
No prior	5.944895391	3443.17263	4.84942E-14	0
Running mean	3.826739421	3742.713532	0.73224538	9
Previous target	3.999596149	3709.848526	0.26775462	9

Figure 39. The mean compression for the separate model fitting process for previous trial and running mean models.



Note.

The model was fitted to each participant on every trial. Fitted net responses were produced by prior weights (empirical) which minimised the SSe. For these separate fits the prior weight was varied for random walk trials and fully random trial separately, then plotted here on the same figure. The compression from the fitted net values was calculated and averaged as it was for Figure 35.

4.6 Discussion

The current research aimed to reveal how Bayesian priors were acquired and updated during spatial localisation. Firstly, empirical support for Bayesian integration was provided because participants weighted the prior more heavily when sensory likelihoods displayed increasing noise. This was demonstrated in the cue SD varying conditions because as cue SD increased from 0.01 to 0.2, so did the mean compression magnitude. According to a Bayesian perspective, the increasing compression magnitude produced was caused by the more variable likelihoods (larger cue SD's). Increased weighting of the prior was needed to increase reliability of estimates in response to more variable sensory likelihoods.

Similarly, increasing the random walk of the target (walk SD) from 0.01 to 0.2, led to a decrease in mean compression magnitude, but this proportional relationship between walk SD and prior weighting was not as systematic as the relationship between the cue SD and the prior weighting in the cue SD varying conditions. The variability in the random walk, in contrast to the variability of the cue, was defined by the SD of the target positions Gaussian, and not the SD of the overall prior (target) distribution. This means a direct relationship between the walk SD and compression magnitude would have been unlikely because the walk SD did not provide an exact one to one mapping of the variability of the prior distribution.

The difference in compression between the fully random and random walk conditions had the potential to reveal features of the prior updating strategy used by participants. Irrespective of whether the cue SD or the walk SD was varying, a decrease in the compression magnitude in the random walk conditions, relative to the fully random conditions would have indicated the use of a previous target model. This is because the previous target model is a very well-suited strategy for random walk (gradually changing) stimuli and therefore this model would predict a reduction in bias for this type of stimulus order. Some reduction was shown, albeit small, visually in Figure 35, however a main effect of stimulus order was not demonstrated in either ANOVA. Additionally, the previous target model used in the current research also predicted a decrease in compression that was larger than the participants data (see Figure 38). Therefore, this did not provide sufficient evidence that a previous target model was used across both conditions, as it was in Glasauer and Shi (2019;2021b).

However, Glasauer and Shi (2019;2021b)'s finding of reduced compression for random walk relied on the assumption that participants would be using the same fixed (Kalman filter) strategy for both stimulus orders. The current research aimed to provide a closer inspection of the Bayesian prior updating strategy by fitting individual trial data. This process provided reason to suggest participants might be flexible in their strategy according to the stimulus order. This was done by fitting 3 possible models to participants data in all conditions (no prior, running mean, previous target).

Firstly, the models were fitted to all conditions (simultaneous fits) and there were large differences between the data and the models. Irrespective of whether the cue SD varied or the walk SD varied, the previous target model showed a much greater difference in bias between the random walk and the fully random stimuli compared to the data. This difference (reduction) was not as large in the running mean model and was therefore more similar to the data. However, the models provided a relatively poor fit when they were fitted to both stimulus sequences simultaneously. To improve upon this, the models were fitted separately (prior weights varied for one sequence at a time). This provided better fits to the data and resulted in less of a difference between walk and full sequences for both models.

Akaike weights were used to quantify the best fitting models for both simultaneous and separate fitting processes. Overall, the model fitting process suggested that 1) Bayesian models were used because the no prior model was never the best fitting model and 2) that the best fitting model, when fitted separately to each condition, was the previous target model for the random walk trials and the running mean for the fully random trials. Therefore, these findings implied that the prior updating strategy was influenced by the stimulus features to some extent.

In contrast with the finding of a fixed distribution estimation strategy in the temporal domain chapter of this thesis, the spatial prior updating strategies used here showed some flexibility in when they were recruited according to their suitability for each stimulus environment. Spatial features of a stimulus may be more able to cue the 'switch' to an iterative strategy because it is common in nature for the spatial location of an object to be autocorrelated as it moves. Whereas temporal magnitudes are often, but not always uncorrelated.

Even when participants were provided with the stimulus sequence features that encouraged switching to the model that was most appropriate for the environment, this does not necessarily mean that priors weights were optimal. Despite the prior weighting decreasing or increasing as expected according to the respective likelihood and prior variability, the prior was consistently underweighted and the likelihood overweighted relative to optimal weighting. This was in line with previous research (Bejjanki et al., 2016; Kiryakova et al., 2020). Kiryakova and colleagues (2020) suggested that a potential reason for this may be that participants are only considering internal noise (the error relating to their own previous estimates) and not the external noise associated with the sensory cue. To investigate this possibility, they conducted a control experiment where participants were asked to find centroids of dot clouds and in this case the internally generated noise purely from their own responses could be calculated because they were no longer having to consider the bias of the prior. After some adapted modelling (the external likelihood SD was replaced with the participants likelihood SD from the control experiment), these data showed a reduced weighting of the sensory cue compared to the original data, implying that this could explain why participants overweighted the likelihood.

Overall, the current experiment replicated the findings of Bayesian integration during spatial localisation (Bejjanki et al., 2016; Berniker et al., 2010; Kiryakova et al., 2020), because as the cue coin location (likelihood) became more unreliable, the target coin location (prior) was weighted to a greater extent and led to increased compression magnitude as a result. In addition to this highly well documented finding, the current research sought to extend on this and found a change in participants model strategy (fully random or random walk) as demonstrated by Bayesian modelling procedures. The current research has uniquely showed that certain participants may be able to switch their prior updating strategy according to the features of the stimulus sequence. This will require further exploration as this was the first demonstration of Bayesian model switching in the spatial domain.

5. The development of Bayesian prior strategies from childhood to adulthood in the spatial localisation task.

5.1 Introduction

The weighted combination of sensory likelihoods and priors act to improve the precision of estimates but is often accompanied by biases of participant's estimates towards the prior mean. This has been shown in a number of different tasks including temporal reproduction (Glasauer & Shi, 2019; Jazayeri & Shadlen, 2010), motion perception (Weiss et al., 2002) and path integration (Petzschner & Glasauer, 2011). Recent research used the spatial localisation task to pose the question as to whether the ability to form and weight priors is innate or, whether it develops as we age. Chambers and colleagues (2018) adapted the 'Catch the Coin' task, to show that children aged six to eight years were only able to learn the mean of the prior, whereas the older children and the adults were also able to learn the variance of the prior (Chambers et al., 2018). This gives some indication that perceiving in a Bayesian way is learned during childhood. A concluding point of their research detailed the need to study the conditions under which we might be Bayesian as to better understand how prior updating mechanisms may operate in different environments. The authors further emphasised that prior learning could occur in response to changes in long-term (developmental) and short-term (experimental) contexts. The introduction to this chapter will outline the different prior learning shown in adults and then infer from research carried out with children how they may be similar or dissimilar in their learning.

Bayesian model strategies in adults and children: static

In order to answer the question of whether children learn Bayesian priors (and when), we must first look to how adult perception has been interpreted through the use of Bayesian models. There are likely at least two possibilities of how a Bayesian prior may be learned and updated over time. The first being a static strategy, where the prior mean is learned by calculating the summary statistics (mean and SD) of all the stimuli experienced so far (Glasauer & Shi, 2021a). As described in Section 1.6, initial evidence supports the use of static models in adults (Jazayeri & Shadlen, 2010). Furthermore, research with children also provides evidence which suggests use of some sort of distribution learning model as outlined in Section 1.9.2 (Karaminis et al., 2016).

Bayesian updating strategies in adults and children: previous trial.

More recently, it has been proposed that priors may be updated, not by learning the mean of the stimuli distribution, but by using information from the stimulus presented in the previous trial. As described earlier in this thesis, Glasauer and Shi (2019;2021) aimed to explore the influence of the experimental paradigm (fully randomised presentation of stimuli) on central tendency biases. This reduction in central tendency bias in random walk conditions was found by Glasauer and Shi (2019) and as a result they claimed that the trial-to-trial prior formation is likely driven by a Kalman filter.

Hallez and colleagues (2019) carried out research with children and used a multiple regression like analysis to explore the influence of the previous trial. The youngest children (5 years) showed the strongest reliance upon the previous trial. Whereas older children (7 years of age) did not show a clear pattern of diminishing reliance on the previous subjective durations to the same extent. This strong reliance on previous trials, as shown by the five-year-old children, appears most similar to the operations of the previous trial, (or a Kalman filter like) model where prior learning is focused on the most recent changes to the stimuli (Glasauer & Shi 2019). Whereas the behaviour of children beyond this age was more ambiguous, as their reliance on the previous trial was not as strong. Additionally, we cannot claim Bayesian origins to these strategies at this stage because there was an absence of computational modelling processes. The current research aimed to see which Bayesian model provided the best fit for children and, to understand the mechanistic processes used by children 6 years old and upwards.

Current research

It is likely that there is a transition from childhood into adult Bayesian strategies but exactly how children may use Bayesian strategies as they age has not been explored. Research conducted by Chambers and colleagues (2018) suggested that whilst children of all ages could learn the prior mean, they could only learn the prior variability (SD) once they were 9 years old. This result suggested a developmental trajectory for prior weighting because the competency in weighting a prior according to its uncertainty was eventually attained by age 9. However, the current research aimed to explore whether the same would be true of the children's ability to switch their prior learning strategy according to stimulus sequence cues. The current research used both random walk and fully random sequences to test whether children were sensitive to cues in stimuli that encouraged different prior acquisition strategies.

From the spatial research in Chapter 4, there is evidence to suggest that adults were able to switch their model according to condition. This was shown by the finding that the largest Akaike weights for fully random conditions were produced by running mean strategies, but the Akaike weights were largest in the random walk conditions for the previous trial model. The current research aims to investigate 1) whether children show the same model as adults when analysing the best model with fixed parameters across all conditions or whether a different model provides a better fit, and whether 2) children can be flexible in their model according to condition as the adults were. It is also possible that children of all ages could show a fixed approach, without the ability to switch models according to stimulus cues. If this were the case, the best fitting model (even when fitted separately to condition) would be the same irrespective of the stimulus order in that condition.

There are reasons to believe that children may learn priors differently due to certain characteristics they display during temporal estimation. Children have shown increased variability in temporal discrimination and in a series of experiments this variability decreased with age (Droit-Volet, 2003; Droit-Volet et al., 2003; Zélanti & Droit-Volet, 2011). In a more recent experiment using temporal discrimination and reproduction, a decreasing variability with age was also shown whereby children reached adult levels of variability at around 10 years old (Karaminis et al., 2019). In this same study, they also fitted a linear regression for the physical duration as a function of reproduced duration to analyse the extent to which reproduced durations were biased towards the mean of the physical durations. Increased bias was found in the youngest children, and this decreased with age (Karaminis et al., 2019).

As a result, the current research also aimed to replicate the 1) increased variability and 2) the increased compression (bias towards the prior) in children relative to adults. Specifically, children aged 6 to 11 in order to observe the developmental trajectory and possible distinction in prior learning between the ages of 8 and 9 years old. Additionally, the different Bayesian model strategies will be fitted to see which models were the most likely to be used by children at these different ages, again aiming to understand how this ability may change at around 9 years old. An adult level for learning prior variance during spatial localisation (Chambers et al., 2018) was shown to be somewhere between 8 and 10 years and, whilst unexplored until this time, it is possible that changing model strategy in response to sequence order may happen at a similar age. Alternatively, children may show a different pattern to adults in the model they chose, with the transition to adult like behaviour happening at an age beyond that of the current sample (12 years and older). It would be surprising if the younger children displayed an adult ability in relation to prior learning behaviour, although it cannot be ruled out because young children have displayed the capability to learn the prior mean in Chambers and colleagues (2018) study.

5.2 Methods

Participants

33 children (17 males and 16 females) were recruited through a public engagement event run by the University of Nottingham. There were 18 six to eight-year-olds, 15 nine to eleven year olds. An experiment room was provided by the event, and they were all assigned a 4-digit ID to anonymise the data. They had no formal diagnoses of developmental conditions, and their vision was either normal or corrected to normal. The adult sample included the same participants, and their data, as in Chapter 4.

Apparatus and equipment

The research was conducted using a Psychopy experiment. The experiment was designed to be identical to the adults but with the addition of more interactive instructions prior to the start of the task (see Stimuli section for more details). The experiment was run on a HP laptop and was held in a large university teaching room as part of the Summer Scientist Week event.

Stimuli

See Chapter 4 for stimulus details that remained the same across the adults and the children's' task. The main focus of this section will be how the instructions were presented to aid children's' understanding of the experiment.

The instructions were designed to help children understand the significance of the two coins, as there was a concern during piloting that the child felt the aim of the experiment was to remember exactly where the cue coin was. Therefore, it was necessary to get across to the children that the yellow cue coin would be located close to the red coin, but that the yellow coin could vary slightly in position each time. It is possible that children may struggle to imagine what a noisy cue such as this would look like, when instructed in words as opposed to when having concrete experience with said cue. It has been shown that children find abstract words more challenging to read than concrete words. According to dual coding theory this could be because both types of words contain a verbal code, however concrete words also contain an imagery code that is likely produced through physical sensory experience with the word (Caramelli, Setti, & Donatella, 2004).

In addition, the lack of a clear one to one relationship between the coins prior to the task might lead to them feeling unsure what the goal of their task is. As a result of the above reflections, the instructions included some self-paced (the space bar needed to be pressed before the next stimulus appeared on the screen) practice trials to give the children a more concrete experience of where the red and yellow coin might appear prior to starting the experiment. They were not explicitly told the roles of the red and yellow coins in a Bayesian context, but they were given the opportunity to experience the concrete elements of the noisy relationship between the two coins.

The initial instruction screens involved laying out what was going to happen in a similar way to adults, although with reference to its pirate and treasure coin theme. Subsequently, the more involved instructions were presented in the form of self-paced and labelled practice trials. Each of the self-paced practice trials had 3 main components: one practice trial showing them that the red coin could appear to the left of the yellow coin, another showing them that the red coin could be to the right of the yellow coin and finally a trial showing them how the yellow and red coin could be almost exactly in the same place. This was done to allow them to know before-hand that there were several possibilities and, whilst the yellow coin did not consistently reveal the exact location of the red coin, the relationship between them existed all the same (see Figure 40). Although no details about the extent of help the yellow coin would provide from trial to trial were given. The details of the self-paced practice trials will be elaborated on the in the next paragraph.

The first interactive instruction page began to detail the relationship between the coins and did so by telling children that a fixation, followed by a yellow coin would appear at the start of all trials. The purpose of the yellow cue coin was explained by telling the children that it would sometimes help them find the hidden location of the red coin, but not always. On this same screen there was some text to describe where the yellow coin could be placed relative to the hidden location of the red coin (for example, to the right) and then they were allowed to move the net around that page to practice looking for the red coin.

In contrast to the actual experiment the yellow coin remained on the screen for the duration of their search to help the children observe the difference between the yellow and red coin’s position. This was preferable to asking them to retain the yellow coin’s position in memory and focus on the instructions simultaneously. Once they pressed the space bar, the hidden location of the red coin was revealed and, again the yellow coin remained on screen so they could see where the coins were in relation to one another. Once they had received that feedback in the form of the red coin (the net would still turn green), they pressed space bar to continue through all the three self-paced practice trials that were set up in a similar fashion (see Figure 40). As part of this self-paced approach, the yellow coin remained on the screen in all but the final instruction screen, so the children could observe what was going to happen in relation to the red coin and the yellow coin positions. After this time, five experiment paced practice trials were used to allow the children to experience the difference in stimulus presentation in the experiment relative to what they had seen during the self-paced practice. This meant in the first case they were able to practice understanding, what would happen and in the following case they were becoming accustomed to how fast the experiment would be. For a comparison to the instructions the adults received see Chapter 4.

Figure 40. A schematic showing how the self-paced practice trials operated as an extension of the instructions.

Here is what the yellow coin will look like.

Your job is to move the net to have a guess where the next red COULD be (move the mouse) and press space when you are happy with your guess.

Remember the yellow coin is not always helpful!

The red coin could fall to the right of the yellow coin, like this…

Press space to continue

Or the RED coin might appear to the left of the cue coin. Have another go to see what that will look like.

Press space to continue

Finally,the red coin could help you and be exactly on the same place as the yellow coin. Have another go to see what that will look like.

As seen here,the red coin could be exactly on the same place as the yellow coin. But watch out as this won't happen every time! Press space to continue.

Note. From the top right panel onwards, the red coins position is revealed to potentially be to the right of, the left of or in the same location as, the yellow coin. The purpose of this was to explain to the children some of the potential outcomes of the noisy relationship between the yellow and red coin.

Procedure

The main procedure was the same as the procedure for the adults as described in the spatial localisation task. However only a single value of cue SD (variability of the cue coin position) and single walk SD (variability of the random walk) was used. Whereas the adults experienced additional variations of the cue SD and walk SD values. This was done to keep the task length (20 minutes) suitable for the attentional capabilities of the youngest children. Both the adult and child version of the task had a score, which was not used for analysis but used to reduce boredom given the simple task procedure.

The procedure began with the instruction screen and the children were helped with the computer responses where necessary for younger individuals. Subsequently a block of either random walk trials or fully random trials with walk and cue SD both equal to 0.05 was presented (the following paragraphs will outline this with more detail). The order of which the participants viewed the two different stimulus blocks was counter balanced.

In contrast to the adult task, the block of trials was split in half with an image telling them they had collected half of the treasure. This meant that children filled two treasure chests and due to both having a short break screen at the half-way point this meant the task appeared to the children as broken into 4 shorter tasks. There were 68 trials in the random walk block and 68 in the fully random block. As a result, there were 34 trials in each quarter separated by a break and 136 trials in total. Overall, each of the conditions took approximately 10 minutes, producing a total time of 20 minutes to 25 minutes including breaks.

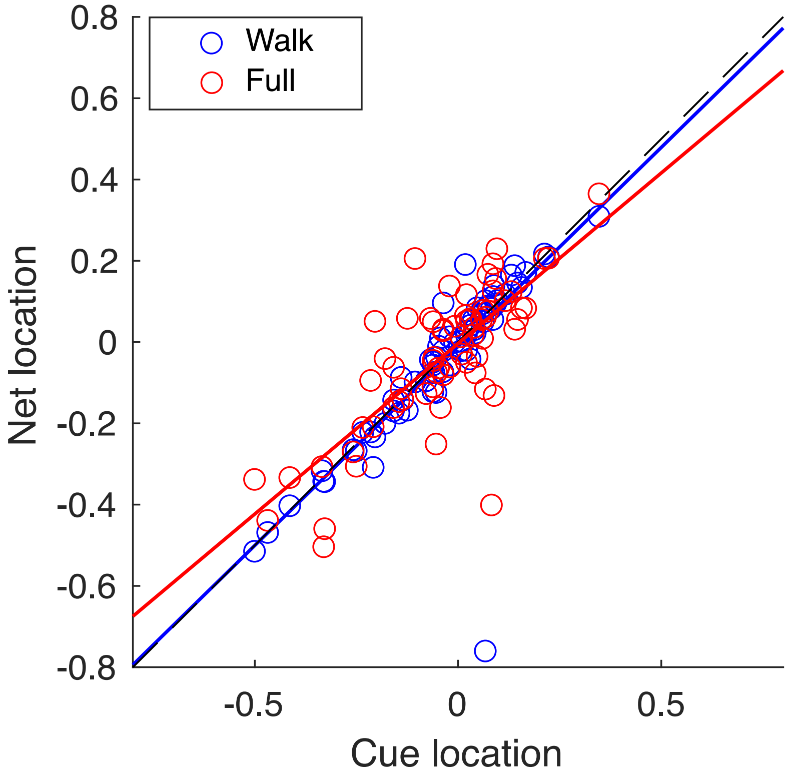
The choice to select these values for the task was decided after analysing the results of the adult data. One condition was selected to keep the task under 20 minutes for the children. This condition was selected because based on the adults' data we were looking for a noticeable difference between the random walk and fully random condition. The differences in compression between conditions and groups is more easily analysed if we can be more confident that there will be a difference between conditions (especially given that children demonstrate more sensory noise than adults (Droit Volet, 2013). This large difference in mean compression was found in cue SD = 0.15 and walk SD = 0.05 (see Appendix G). Additionally, the standard deviation of the cue coin needed to be small enough so that children could begin to see the relationship between the two coins. As a result, the cue SD = 0.05 condition was selected.

The adults experienced 5 different walk SDs (one fully random and one random walk) and 5 different cue SDs (one fully random and one random walk), leading 4 different phases. This was done to demonstrate a systematic effect in how the bias towards the prior changed, before looking into developmental differences in the current chapter.

5.3 Results

Estimates of target coin locations were collected via a net response which was contingent with participants' mouse placement. The cue coin locations (as extracted from the Gaussian curve where the target coin position is the mean) were plotted as a function of the net location responses as predicted by a single participant (see Figure 41). The random walk conditions were shown in blue, and the fully random conditions were shown in red. A linear regression line of the same colour was fitted through the data points for each condition. A linear regression of cue positions as a function of responses (net position) was fitted. This is because the amount of bias can be measured by the amount of bias towards prior (target location mean; see Figure 41). For both conditions, the linear regression can visually demonstrate the amount of compression towards the target (prior mean). This is because, if the net were to be placed on the cue coin every time, then there would be no compression and the slope would have a gradient of 1. If the net location were to be biased towards the target mean, then the gradient of the fitted line would move away from 1 and towards zero. As a result, the size of the slope indicates how much bias there was towards the prior mean such that a smaller gradient value (closer to 0 would mean more compression). However, to make the measure of bias more intuitive 1-b was used because then a larger value of 1-b would mean more bias towards the mean.

Figure 41. The unfilled data points correspond to the net locations plotted as function of cue locations for a single participant.



Note. The net locations correspond to where participants thought the subsequent target location would be and as result where they placed their net at the end of each trial. The cue locations were drawn from a Gaussian distribution with a mean equal to that of the target coin location on that trial and a standard deviation of 0.05. The red circles indicate the fully random condition and the blue show the random walk conditions. The black dotted line represents complete reliance on the cue coin for estimating the target coin's location.

The word compression was used to describe how the linear regression fit line appeared flattened. Compression magnitudes (1-b, as described in the previous paragraph) were calculated for each individual participant, in each condition. These values were then averaged and plotted against the standard deviation of the cue and the standard deviation of the random walk step respectively to see how bias changed as the variability of these stimulus properties was altered. Random walk and fully random conditions for each cue and walk SD were plotted on the same figure in order to provide a comparison for the different amount of compression shown for the two stimulus sequences.

The data taken from a single participant was plotted with a linear regression (see Figure 41). The data from this participant showed that random walk conditions produced very little compression towards the target location and almost complete reliance on the cue coin. This is demonstrated by the random walk linear regression line being almost exactly in line with the dotted line, where the dotted line represents the net being placed on the cue coin every time. However, for the fully random conditions, there was a clear compressive bias away from the dotted line towards the prior mean (mean of the target locations).

Group data bias: compression magnitude.

The compression magnitude was averaged across the two groups of children (6-8 years and 911 years) and the group of adults. The bar plots in Figure 42 show the different amounts of compression in each age group. The data for the two child groups was also plotted on a continuous scale to assess the extent to which a gradual change in compression magnitude exists between the ages of 7 and 9.

The amount of compression decreased with age, such that for the fully random conditions the younger children showed the largest compression (M = 0.387, SD=0.264), followed by the older children (M=0.180, SD=0.131), followed by the adults (M=0.107, SD=0.118). This same pattern was seen for the random walk conditions where the younger children (M=0.346, SD=0.208) showed the largest compression magnitude, followed by the older children (M=0.135, SD=0.181), followed by the adults (M=0.056, SD=0.049).

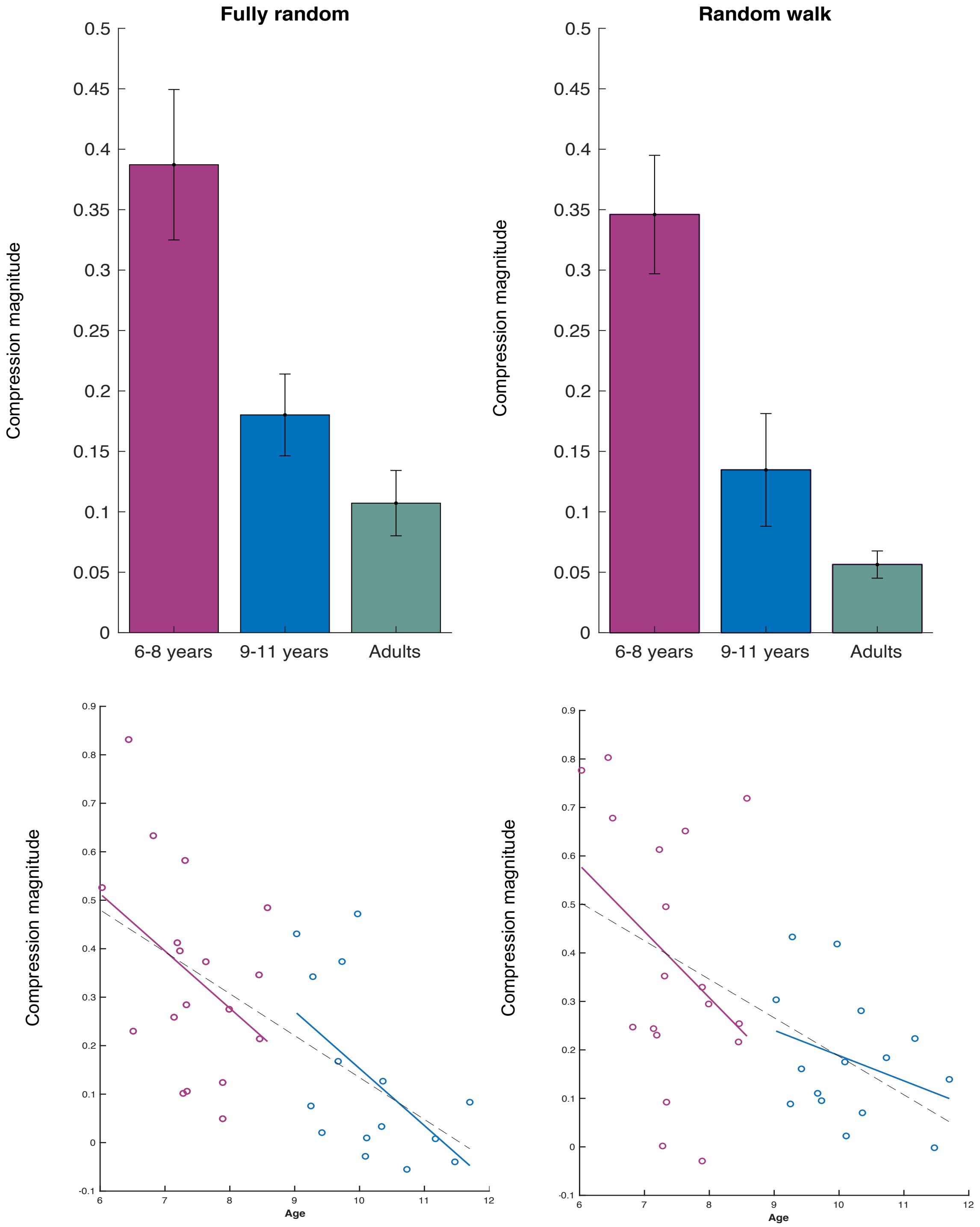
Additionally, fully random conditions showed visibly larger compression than random walk conditions (apart from in the 6-8 year olds where this difference was small compared to the other age groups). Younger children (Marginal mean = 0.367) showed more compression than (Marginal mean = 0.157) in terms of the amount of compression they displayed, irrespective of condition. Additionally, younger children

(Marginal mean = 0.367) also showed more compression than the adults (Marginal mean = 0.082) again, irrespective of condition (see Figure 42). As a result, the differences between the older and younger child group when presented as categorical data appeared to suggest a negative relationship between age and compression magnitude.

However, linear regressions were fitted to the data to assess whether the relationship was present within the two age groups (as well as between them). In the fully random trials, the 6-8 year olds showed a negative relationship between age and compression magnitude (R squared=0.163, F(2,16)=3.12, p=0.0996 was found. For these trial types, a negative relationship was also found for the 9-11 year old children (R squared = 0.286, F(2,13)=5.2, p=0.04). Additionally, the linear regression fitted across all child groups found a negative relationship between (R squared=0.381 , F(2,33)=19, p<0.001)

In contrast, in the random walk trials the younger children's' data showed that age was not able to predict compression magnitude (R squared = 0.135, F(2,16)= 2.5, p = 0.133. This was also the case in the In the older children's data. Age was not predictive of compression magnitude (R squared = 0.106, F(2,13)=1.54 p=0.236). However, for the linear regression fitted across children of all ages a negative relationship between age and compression magnitude was found (R squared=0.282, F(2,33)=12.2, p=0.0015)

Figure 42. The amount of compression magnitude towards the target mean in random walk and fully random conditions for all age groups.



Note. The compression magnitude for each participant was calculated using the steepness of the fitted linear regression line seen in Figure 41. More specifically, the compression magnitude was equal to 1-b, where b was the gradient of the slope. Therefore, fully veridical responses would have produced a gradient of 1 and a compression magnitude of 0.

For the categorical data the average compression magnitude was then calculated for each individual and averaged across each group to produce the bar heights. The error bars represent the standard error of the mean. For the continuous data, the precise age of the children (participants from ages 6 to 8 and 9 to 11) was plotted against their corresponding compression magnitude. A linear regression line was fitted across all ages (dotted line) and within each age group (filled lines).

Mixed factors ANOVA on compression

A 3 (age: 6-8 years, 9-11 years or adults) x 2 (randomisation: full or walk) mixed factors ANOVA (assuming unequal variances) found a significant main effect of age on compression was found F (1,49) = 17.279, p<0.01, eta squared 0.348). A Levene's test indicated that the assumption of homogenity was violated for the full (F(2,49) = 10.817, P<0.001) and the random walk (F(2,49) = 10.211, P<0.001) conditions.

A post hoc Tukey test revealed that there were significant differences between the 6-8 year olds and 9-11 year olds (p=0.01) and the 6-8 year olds and the adults (P<0.001).However, there were no significant differences between the older child group (9-11 year olds) and the adults. From the 3x2 mixed ANOVA it was also shown that there was no main effect of randomisation on compression F(1,49) = 3.984, p=0.051, eta squared=0.012) and no interaction between randomisation and age in terms of the amount of compression shown, F(2,49)=0.016, p=0.984, eta squared=0.000094).

Group data variability: Standard deviation of residual error

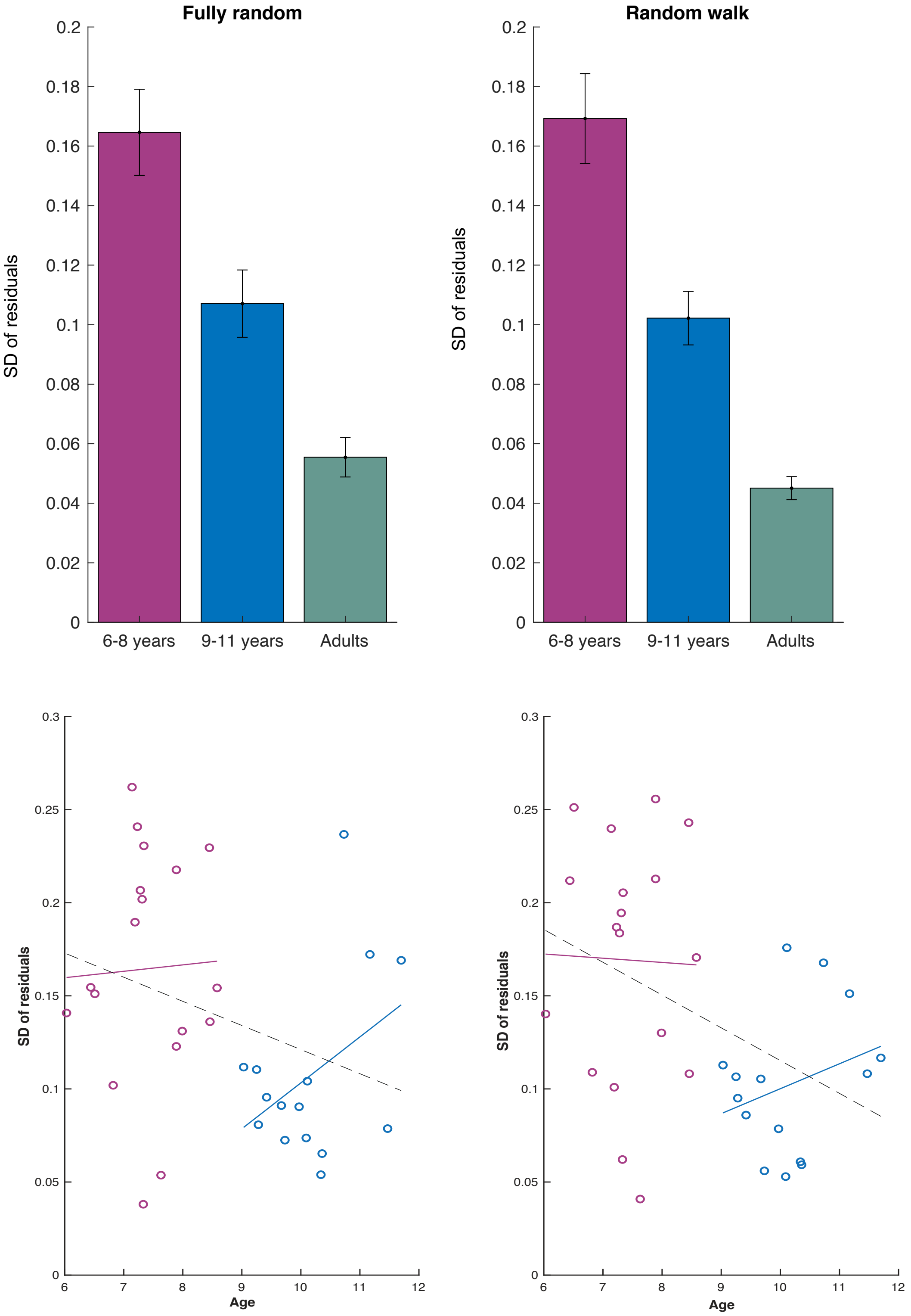
The variability of participants was measured via the standard deviation of the residual error across all trials in each condition, for each participant. The residual error was the target position subtracted from the predicted target location (as predicted from the linear regression) for each participant. This error was averaged across each age group and plotted in Figure 43 as categorical data to highlight the different levels of compression across the age groups.

In the fully random conditions, the SD of residual error was largest for the 6-8 year olds (M =0.165, SD=0.063) and showed a decrease as the age of the participants increased (see Figure 46). The SD of residual error was reduced relative to the 6-8 year olds for the 9-11 year olds (M=0.107, SD=0.049) and then reduced further for the adults (M=0.055, SD=0.029). The same pattern emerged for the random walk conditions with the largest SD of residual error being associated with 6-8 year olds (M =0.169, SD=0.066), the next largest being the 9-11 year olds (M= 0.102, SD=0.039) and the smallest being associated with the adults (M=0.045, SD=0.017).

The data for the two child groups was also plotted on a continuous scale to assess the extent to which a gradual change exists between the ages of 7 and 9. Linear regression lines were fitted to see whether, within each age group, age was predictive of variability (SD of residuals). For the full random trials, the 6-8 year olds showed a relationship between age and compression magnitude (R squared =0.0015, F(2,16)=0.0242, p=0.878) was not found. It was also true that for the 9-11 children a relationship between age and SD of residuals was not found (R squared =0.168, F(2,13)= 2.63, p=0.129). The linear regression fitted across both child groups also found that age was not able to predict SD of residuals (R squared = 0.104 F(2,33)=3.58 p=0.068).

In the random walk conditions, for the children aged 6 to 8 a relationship between age and SD of residuals was not found (R squared=0.0006, F(2,16)=0.007, p=0.923). For the children aged 9-11 years, a relationship was also not found R squared=0.079, F(2,16)=1.11, p=0.312.The linear regression fitted across all children demonstrated a relationship between age and variability (R squared=0.187 F(2,33)=7.12 p=0.012).

Figure 43. A bar chart displaying the standard deviation (SD) of the residual error in random walk and fully random conditions for all age groups.



Note. The residual errors were calculated by taking the difference between individual responses and the response as predicted by the regression line. Then the standard deviation of those residual errors was calculated for each participant and averaged across each condition and each age group to form the bar heights. The error bars represent the standard error of the mean.

For the categorical data the SD of residuals averaged across each age group was used to form the bar heights. The error bars represent the standard error of the mean. For the continuous data, the SD of residual errors for each child were plotted against their precise age. Linear regression lines were fitted across all ages and within each age group.

Mixed factors ANOVA on variability

A 3 (age: 6-8 years, 9-11 years, or adults) x 2 (randomisation: full or walk) ANOVA was conducted on the SD of the residual error and a significant main effect of age group was found F(2,49) = 33.869, p<0.001, eta squared = 0.536**.** However, there was no main effect of randomisation F(1,49) = 0.463, p=0.499, eta squared = 0.00055 and no significant interaction F(2,49)=0.762, p=0.472, eta squared = 0.002). A Levene's test demonstrated unequal equal variances of the groups for the full (F(2,49) = 6.118, p =0.004) and the walk (F(2,49) = 10.304, p < 0.001) conditions.

Post hoc Tukey tests carried out on the main effect of age highlighted where the differences between the groups were. The 9-11 year olds were significantly different from the adults (p=0.02) and the 6-8 year olds were also significantly different from the adults (p<0.01). The Tukey test also showed a significant difference between the 6-8 year olds and the 9-11 year olds (p<0.01). Taken together, the adults displayed much less variability in their responses compared to both groups of children (see Figure 43).

Model fitting

The same models were fitted here as those described in Chapter 4. As described in Chapter 2 and 4, the lowest SSE will reveal the best fitting model and highest Akaike weights will reveal most likely model. See Appendix N for individual Akaike weights and see 2.2.3 for the formulas used to produce Akaike weights. The most likely model when all conditions were fitted simultaneously to all trial types was the previous target model for all age groups (see Table 4). However, when fitted separately differences appeared between the adults and the two child groups. The adults showed the best fitting model to be the running mean model during fully random trials and the best fitting model to be the previous target model for the random walk trials. The oldest children, however, showed the previous target model to be the best fit for random walk conditions only (see Table 6 for Akaike weights) and fully random conditions only (see Table 5). The younger child group showed a difference according to the stimulus sequence because the best fitting model for random walk trials was the previous target (see Table 6) whereas it was the running mean model for the fully random conditions (see Table 5).

Table 4. The Akaike weights mean averages when the models were fitted to all conditions.

Model	All children	6-8 years	9-11 years	Adults
No prior	0.147744309	0.130851127	0.168016128	1.30496E-36
Previous target	0.520619515	0.499331903	0.54616465	0.587935857
Running mean	0.331636175	0.36981697	0.285819222	0.412064143

Table 5. The Akaike weights mean averages when the models were fitted to fully random conditions only.

Model	All children	6-8 years	9-11 years	Adults
No prior	0.14991317	0.139031443	0.162971243	4.84942E-14
Previous target	0.467215937	0.394553669	0.554410659	0.26775462
Running mean	0.382870893	0.466414888	0.282618099	0.73224538

Table 6. The Akaike weights Mean averages when the models were fitted to random walk conditions only.

Model	All children	6-8	9-11	Adults
No prior	0.301559731	0.192468163	0.296134853	0.040953126
Previous target	0.490221481	0.503158093	0.38700282	0.832806247
Running mean	0.208218788	0.304373744	0.316862328	0.126240627

5.4 Discussion

Compression magnitude (bias)

Previous research has demonstrated that children display increased biases towards the mean of the stimulus distribution, coupled with the increased amount of noise in their responses (Karaminis et al., 2019). In their research, children showed decreasing amounts of compression towards the mean as their ages increased, producing similar levels of compression to adults at around 10 years of age. In the current research, it was found that there was increased compression in the youngest children compared to the other two age groups. The youngest children (6-8 years) showed significantly more bias than the older children (9-11 years) and the adults. Therefore, an overall decrease in compression magnitude was seen as development progressed. In addition to this trend being shown in the visual appearance of the data, linear regressions suggested a negative relationship between age and compression magnitude in all instances except for within each age group in the case of the random walk trials.

The sizeable reduction in compression seen in 9 to 11 year old children relative to 6 to 8 year olds, implied prior weighting akin to adult levels. This is in contrast to the younger children who have placed a much larger weight on the prior. During piloting, there was a concern that younger children, may show reduced weight on the prior (or potentially cue only models), if they were overly fixated on the yellow coin as it was the first stimulus that grabs their attention. This provided the motivation for using more interactive instructions to aid children's access to the statistics of the task to their best of their ability despite limitations in executive functioning, attention and memory. Therefore, also reducing the possibility of them using cue only models due to misunderstanding the task or misremembering the purely linguistic instructions. As a result of the interactive instructions (slowed down practice trials with extra warning about the uncertainty surrounding the cue), there is the possibility of further increases in compression compared to adults. However, due to the older children and adults showing no significant differences in the amount of compression they displayed. This emphasised that despite the addition of child friendly instructions, compression for the older children was merely visible around the adult level, and not falsely inflated above it.

SD of residual error (variability)

Based on previous research, it was expected that the variability associated with responses would decrease with age, with the youngest children showing the largest amount of variability and more adult like variability being shown around the age of 10 (Droit-Volet, 2003; Goldstone et al., 1958; Goldstone & Smythe, Elizabeth, 1957; Rattat & Droit-Volet, 2001). However, other research has placed a transition into adult perception somewhere between the ages of 7 and 8 (Espinosa-Fernández et al., 2004).

This expected decrease in variability with age was demonstrated in the current research using the amount of SD associated with participants residual errors as a measure of individual variability. Post hoc comparisons showed that all age groups were significantly different from one another. The youngest children (6-8 years) showed the largest variability, closely followed by the older children (9-11 years) with the adults showing very small amounts of individual variability. The individual variability displayed by children from both age groups did not reach a level equivalent to adults. It could be the case that tasks such as temporal discrimination where adult levels of variability may be seen between the ages of 7 and 8 (Droit-Volet & Wearden, 2001), for tasks where prior learning is required for the task (Karaminis et al., 2019), variability may not reach an adult level until at least 10 years.

However, a possible reason for not seeing an adult level of variability in the older child group could have been due to the larger number of trials experienced by the adults. Although, motor error that is, the variability in where the net is placed across all trials, is likely to be a large source of error for children (Yan, Thomas, Thomas, & Stelmach, 2000) and this type of variability is likely to be heavily reduced in adults relative to children because motor noise reduces into adulthood (Takahashi et al., 2003). This type of error may not be greatly impacted by the number of trials that children do, as they might have less reserved (and more sporadic) estimates than the adults. On the other hand, the type of variability that would be impacted by practice is the variability of prior representations. It can be seen from the compression magnitude data that there was no significant difference in the compression magnitude between the older children and the adult group despite the adults completing a larger number of trials. As a result, the reduced amount of practice cannot be claimed as having influenced the variability of the prior representations that were learned from the stimuli. In a purely practical sense, especially for the younger children, an increased number of trials would not have been feasible due to the 20 minute task already testing their attention span.

It may be the case that the nature of the relationship between age and variability was gradual as children aged. The visual features of the data suggested a gradual and negative relationship between children of all ages and variability. However, this relationship was only significant when calculated across all the children (aged 6-11) and not within each age group.

Model fitting

Despite clear developmental trajectories being visible for both compression magnitude and variability of responses, the same cannot be said for the Bayesian prior acquisition strategies. Bayesian model fitting was also carried out to understand how children might update their priors in this spatial localisation task. In line with the model fitting conducted with the adult's spatial localisation data, when fitting all conditions, the previous target model was the most successful for children in both age groups. The largest Akaike weight was produced for this model implying that it is the most likely model being used by participants. Therefore, this appeared to be the best fitting model for this task for all age groups. As a result, if a fixed model strategy were to be assumed across the two different randomisation conditions, adults and children appeared to approach the task in a similar way.

It is possible that model choice differs according to stimulus properties. This was true in the case of the adult participants because, in addition to showing a best fit for the previous target model, adults appeared to change their model when the stimulus cues (randomisation sequence) called for such a change. This is because a previous target model was demonstrated to again be the most likely when fitting only the random walk conditions. This was expected because the random walk sequence is especially suited to a model that focuses on learning according to the most recent changes in the stimuli. However, when the model was fitted to only fully random stimulus conditions, where the running mean model would have provided a more optimal estimate of the next stimulus position, the best fit was indeed provided by the running mean model. This suggests that adults flexibly show a different best fitting model according to the stimulus condition.

It is pertinent to this chapter as to whether the children were able to produce the same flexible response behaviour as adults. In the current research when the random walk conditions only were fitted, the previous target model provided the best fit for children. Therefore, they appeared to show the ability to determine that a previous target strategy is suitable for responding to a random walks sequence. However, when only the fully random conditions were fitted, this is where the children's behaviours appeared to deviate from the adults. When only the fully random conditions were fitted to the older children's' data, they appeared to show the previous target model as the best fit. This means that they favour a fixed previous target model in both conditions and did not change according to the stimulus sequence presented. Whether this is due to lack of detection of the stimulus cues or the conscious choice to maintain the same strategy remains to be seen.

Interestingly, for the youngest children (aged 6-8), the running mean model provided the best fit for the fully random conditions. Although upon further inspection of the counterbalancing it appears that the model the youngest children used was in fact in line with what would be most optimal for the stimulus environment that they had experienced first. The previous target model also provided the best fit for the youngest children when fitting data to all conditions. Therefore, it is doubtful the extent which these young children showed a change in strategy like the adults. These data suggest young children reactively select a strategy that is best suited for the first iteration of the task that they experienced.

Different best fitting models according to the stimulus sequence was not a pattern seen in the older children. For children aged 9 to 11 years the previous target model showing the best fit regardless of which order they completed the task in. So potentially an ability to move past using what worked best at the start of the experiment could develop as children age. In the longer format of the adult's task, the task that was presented first had no impact on this switch, the best fitting model, in line with the flexibility to switch, was the same irrespective of which type of stimuli was presented to them first. However, for the youngest children the order of the conditions did have an influence such that the average Akaike weight was highest (most likely model) for the strategy best suited to the stimulus sequence that was presented to the children first. This was shown through the average weight being higher for the strategy most appropriate for that stimuli sequence (for example, the previous target model and random walk stimuli or the averaging model and the fully random stimuli) were presented first. This was not true for the older children because the previous target model was always the most likely for them irrespective of the order of conditions.

It is important to note that the adults also showed many individual differences in their chosen model with no single model being the most likely for everyone in the group. Often the differences in Akaike weight between the two Bayesian prior acquisition strategies were marginal (and not unanimous which seemed to support Glasauer and Shi (2021a)'s finding that participants show some individual differences in their beliefs about stimuli. It is plausible that different people choose a different strategy due to the relatively small amount of difference in their success in catching the coin (especially in the children's' task). The reward provided by the feedback of the task may have had a different effect on children relative to adults or it may have played a larger role in the selection of their preferred prior updating strategy. It also could have discouraged older children (aged 9-11) from switching as they grew confident in their prior updating strategy by using the reward to guide them.

Taken together, there was a clear developmental trajectory in terms of compression magnitude and response variability as children age in line with previous research. However, the same was not true for the most likely Bayesian model strategy. It seems for this task that the previous model was the most likely (highest Akaike weight) for all ages if a fixed model were assumed across all stimulus conditions. Overall, in terms of flexible model strategies, the older children showed a clear preference for the previous target model irrespective of changes to the stimulus sequence. However, the younger children showed the use of a model that was optimal according to the stimulus sequence they experienced at the start of the experiment. Therefore, it is possible that: either 1) there is not a linear developmental trajectory for flexible strategies and so younger children can be flexible but older children cannot, or 2) the more plausible conclusion from these data, is that flexibility is not possible in childhood, and any flexibility in the younger children was due to marginal differences brought about by whichever condition the younger children experienced first. An analysis of the averages of the two different counter balanced groups suggested that the latter was more likely.

It was also possible that the adults were able to (and detected the need to) switch due to an increased number of trials for their task. Although this is challenging to disentangle from their increased attention span to cope with a larger number of trials in comparison to the children. Conversely, in previous research children have shown limited competency in learning statistical regularities such as learning the mean of the prior (Chambers et al., 2018; Karaminis et al., 2019). As demonstrated by the effort made in the current research to create child friendly instructions, it is important that children are given the chance to perform to the best of their ability if research is to produce valid comparisons between children and adults.

6. Bayesian inference as an explanation of 'Time Order Errors'

6.1 Introduction

The primary focus of this thesis until this point has been estimations of a time interval or, of the spatial position of a single stimulus. Unsurprisingly, biases in magnitude have also been produced through the comparison of two stimuli (Hellström, 1979, 1985). One such bias was termed the Time Order Error (TOE) by Fechner (1860), who differentiated between negative and positive TOEs according to the relative under and overestimation of the first stimulus. Negative TOEs involve underestimation of the first stimulus relative to the second stimulus and positive TOEs occurred when the first stimulus is overestimated relative to the second (Fechner 1860, as cited in Hellström, 1985). This bias has been demonstrated for a range of relative magnitude judgements: timing (Dyjas, Bausenhart, & Ulrich, 2012), brightness and size (Patching, Englund, & Hellström, 2012) and weight estimations (Hellström, 2000). TOEs of this form are sometimes referred to as type-A effects (Hellström, 1985), to distinguish them from type-B effects, where difference lumens (DL; the smallest amount of difference needed for subjects to perceive it) are found to be larger when comparison stimuli are presented before the standard (CoSt) than when they are presented after the standard (StCo; Hellström, Patching, & Rammsayer, 2020; Kups, 2021).

Early models of time order error

Attempts to explain biases when comparing the relative value of two stimuli have ignited much debate (Schab & Crowder, 1988).The earliest explanation of the processes involved in comparing the value of two stimuli, namely the 'Difference

Model', proposed that perceived magnitudes were only dependent on the physical size of each stimulus and independent of factors such as stimulus order (Hellström et al., 2020). Research demonstrating the existence of TOE (Fechner,1860) casted doubt on the 'Difference Model' as a theory that could explain the process of comparing two stimuli. The 'Difference Model' was flawed because it did not consider that the two stimuli could be treated differently according to the order in which they were presented, which has since been shown to be highly relevant. As a result, more recent models have aimed to quantitatively explain the existence of TOE biases.

Michels and Helson (1954) produced the first quantitative theory of the TOE. Prior to this, it was assumed that the physical magnitude of the first stimulus was compared to the physical magnitude of the second. However, Michels and Helson (1954) found the impact of the first stimulus to be larger than the impact of the second (Michels & Helson, 1954). As a result, they proposed that the second stimulus was compared to the first stimulus plus the Adaptation Level (the weighted geometric mean of previously experienced stimuli, with their weights being set according to their recency). Bias produced from stimuli comparison has also been explained by the Internal Reference Model (Bausenhart, Bratzke, & Ulrich, 2016; Dyjas et al., 2012). This more recent model maintained the comparison was not between the physical size of the two stimuli being compared. The second stimulus was instead being compared to an internal reference (IR), which is a weighted combination of the running mean of first stimulus of the pair and the IR of the previous target (Dyjas et al., 2012).

Sensation weighting model

The sensation weighting (SW) model proposed that, similarly to previous models, it is not the veridical stimulus magnitudes that are compared during discrimination. The key difference for the SW model and previous models is that both stimuli have an assigned weight. The SW explains that it is not the comparison of two stimuli that is occurring, it is the comparison of two weighted compounds. These compounds were described as the subjective magnitude of the stimulus and some reference level or ReL (the ReL is conceptually similar to the adaptation level in Michels and Helson's model; Hellström et al., 2020).

Fading trace accounts

Fading trace accounts of TOE propose that the memory of the first stimulus fades and is therefore not perceived as accurately as the second stimulus. Initial interpretations of the TOE bias without explicit reference to a fading trace were mentioned by Fechner (1860) and built upon by Hellström (1979). This explanation was more explicitly described as a fading trace account by Schaab and Crowder (1988). There has been support for the mechanism of a fading trace underlying TOEs such that TOE bias increased as the interstimulus interval increased (Fechner 1860; Wolfe, 1898, as cited in Schab & Crowder, 1988). However, the reverse of this effect has also been found because in another experiment, as the interstimulus interval (ISI) increased, the size of the TOE reduced (Jamieson & Petrusic, 1975).

If the production of a TOE was reliant on the memory trace of the first stimulus fading, there are further predictions that can be made from this. For example, if memory degradation alone were to be responsible for the TOE effect, then this bias would persist on occasions where only the magnitude of stimulus one or stimulus two are estimated. This prediction of the fading trace account was explored by Schaab and Crowder (1988) in the supra second range. They found that, firstly, there was no clear relationship between the ISI and the size of the TOE. Secondly, a time order error effect was found during reproduction such that the same duration was reproduced as longer when it was presented second rather than first (Schab & Crowder, 1988). These results are consistent with the presence of negative TOEs during reproduction of supra-second duration ranges.

Crucially, theoretical understanding of TOEs could also be improved through considering the impact of the fading trace of the first stimulus. As a logical progression of the fading trace account, the Bayesian inference model explains that the memory of the first stimulus is held for longer and so becomes relatively more uncertain than the second stimulus (Ashourian & Loewenstein, 2011; Raviv, Ahissar, & Loewenstein, 2012). Due to the increased uncertainty associated with the first stimulus, the sensory representation of the first stimulus receives less weight during Bayesian integration. As a result, more weight is given to the prior knowledge (the mean of the stimulus distribution) of the first stimulus. The outcome of this process is sometime referred to as a contraction bias because the value of the first stimulus (more uncertain) gravitates towards the mean value of all the first stimuli across previous trials (Ashourian & Loewenstein, 2011). The first interval may be under or over-estimated depending on the relative size of the second interval. This explains the consistent over and underestimation of the first stimulus in positive and negative time order errors respectively, because the perceived value of the first stimulus is pulled towards the mean of the 'first stimulus' distribution

In some cases, this can be beneficial and provide an estimate close to the true value, however if the current stimulus is far from the mean, this could lead to an unhelpful bias to the participants' estimates (see Figure 3, Section 1.1). Ashourian and Loewenstein (2011) tested the merit of the Bayesian model. During a TOE task, they found a shift in the prior mean in response to a shift in the mean of the stimulus distribution. Additionally, they found that when the range of the stimulus was increased, the prior width also increased (Ashourian & Loewenstein, 2011). As a result, Bayesian models showed some merit in explaining contraction biases during TOE.

According to the predictions of the Bayesian model, a reproduction bias in a direction consistent with TOE should still be visible when one of the two stimuli are reproduced. As a result, the increased uncertainty in the first stimulus should lead to an increased central tendency bias for the first stimulus and result in a TOE. Many current Bayesian modelling studies of magnitude estimation have focused on reproduction of a single stimulus. However, given the emphasis that the SW puts on the comparison of two stimuli, it is necessary to understand whether reproductions of the first interval become biased, even when paired with a delay as a replacement of the other stimulus in the pair. This raises the question of whether TOE bias is only present when comparison is carried out, as predicted by the sensation weighting model. Or, alternatively whether the first interval shows biased perception irrespective of whether the two stimuli are compared or reproduced, as could be explained by a Bayesian model. The assumptions of the sensation weighting and Bayesian models will be covered in the following paragraphs in order to outline the predictions of the current research.

Sensation weighting model assumptions

The sensation weighting (SW) model cannot account for a time order error occurring outside of a two-stimulus comparison context. This is because comparison and interaction of the two intervals is central to the framework of the SW model. Although the fading trace account is currently incorporated within the SW model, there are some issues with this. If the fading trace account of the TOE holds true, then the amount of time that passed between the first and the second interval would have been important for driving the bias. As a consequence, a TOE should be produced regardless of whether both stimuli in the pair were presented or whether a delay replaced the first stimulus. The sensation weighting model only provides assumptions for the comparison of two stimuli and therefore would have no explanation for this possibility.

Bayesian model assumptions

A Bayesian model, however, assumes that it is the uncertainty (contained within the fading memory trace) driving the TOE bias. As a result, reproduction of a single interval (using a delay instead of a stimulus in either position) could cause a TOE under the Bayesian framework. If a Bayesian model can provide a sufficient explanation for the bias of intervals during reproductions of a single interval on each trial, this would provide an insight where the SW model cannot. The current research predicted that a time delay in place of D1 would be sufficient to produce the TOE bias during ‘discriminations’ that were inferred from reproduction data. This would strengthen a Bayesian interpretation of the bias.

However, if data from trials where two stimuli are presented and compared prior to asking participants to reproduce the target duration, (combined trials) produced an even larger TOE than regular reproduction, the sensation weighting may provide a stronger explanation of the TOE. In that case, comparison of the two intervals, in addition to memory degradation, would be necessary for a more sizeable TOE.

Current research

As shown in previous research, the existence of TOEs would be expected during the discrimination of two stimuli (Laurie Anne Sapey-Triomphe, Timmermans, et al., 2021). However, the current research additionally aimed to explore whether the TOE bias remains when only the first or the second interval is reproduced by participants. This was made possible by providing participants with a delay instead of the first or second interval. If a TOE bias is shown in this case, then the sensation weighting model cannot explain the appearance of the TOE during reproductions because its current mechanism relies on participants comparing a sequence of two stimuli.

6.2 Methods

Participants

7 participants (5 male and 2 female) with and age range from 25 to 44. All participants had normal or corrected to normal vision.

Stimulus details

The information concerning the screen position and set up were identical to that of the temporal experiments from Chapters 2 and 3. Similarly, the target stimulus itself was a single circular disk, with the same dimensions as the temporal experiments, was used to represent either a D1 or D2 stimulus duration.

The first interval (D1) on each trial was drawn from a log-spaced range (0.3116, 0.4500, 0.6416, 0.8999 and 1.27490 with a geometric mean of 0.64). The first interval presented in the trial will from here on be referred to as D1 intervals. The steps between each interval will be referred to as the D1 step. Additionally, the second interval in all trials will be referred to as D2. In a similar vein, the difference between the current D2 and D1 (the difference between intervals on any particular trial) will be referred to as the D2 step. The size of the D2 step was 0.075 for all participants. This step size was used on the basis of producing an accuracy level of approximately 75% correct after practice trials were completed.

Procedure

The time order error task contained three trial types, one in which participants were instructed to discriminate and this was followed by another trial type where they were asked to reproduce a single duration (either D1 or D2 with a delay replacing the other comparable duration from the discrimination task). Finally, there were combined trials which will be elaborated on in a later section. For all the aforementioned trial types, D1 appeared followed by an interstimulus interval (ISI) of one second. Practice trials were used to introduce the participant to the task and, in the case of the discrimination trials, to assess their accuracy.

Discrimination trials

The discrimination trials involved participants being asked to judge whether the first or the second interval was longer (with the left button indicating that the first interval was longer and the right button indicating the second was longer). The button box described in Chapter 2 was also used in this experimental set up.

Each of the 5 D1's was displayed 20 times to produce 100 trials in a single run. Of the 20 presentations, the D1 was greater than D2 for 10 trials (a D2 step of -0.075) and D1 was smaller than D2 for 10 trials (D2 step of 0.075). Participants completed a minimum of 6 runs of the discrimination trials. The number of runs was increased if participants data was especially noisy.

Reproduction trials

Reproduction trials required participants to reproduce a single time interval. Either the first or the second interval of the stimulus pairs was presented (the same values of D1 and D2 as were used for the discrimination trials). The other interval in the pair was replaced with a delay, creating either a delay before the stimulus interval appeared (when the delay was replacing the second duration in the pair), or after (when the delay was replacing the first duration).

The stimuli during reproduction trials were presented such that they mimicked how the discrimination pairs were presented as closely as possible. However, the trial structure was slightly different as a result of the reproduction of single durations. The fixation cross was used to help participants know when to respond after the delay. After the white fixation was presented at the start of the trial, then one of two possibilities occurred: either a delay followed by D2 or D1 followed by a delay. After both the stimuli and the delay have been presented the fixation cross changes to indicate that participants could respond. The white fixation cross reappeared at the end of the trial and, it turned black as participants reproduced the duration. If they pressed the response button before the end of the trial, the fixation would not turn black, and the response would not be counted. As with discrimination, practice trials were used to help participants to adjust to the response period.

Of the 10 presentations of D1>D2, only the D1 duration (with a delay after) was presented for 5 trials and only the D2(with a delay before) was presented for 5 trials. The 50/50 split was also present for the 10 D1<D2 durations. As a consequence, for each D1 and D2 duration, there were half the number of trials in a run (50 instead of 100). This meant that participants had to carry out a minimum of 12 runs for the reproduction trials to produce an equivalent number of trials for reproduction and discrimination. This produced an overall experiment time of approximately 2 hours and 15 minutes.

Combined trials

The combined sessions were purposely designed such that discrimination and reproduction could take place within the context of a single trial. After the procedure above, participants returned for an additional 1 hour session where they were presented with the same values of D1 and D2 but this time they were asked to reproduce the duration that they felt was longer. They were required to press the left button if they thought the first duration (D1) was longer, and the right button if they thought the second duration (D2) was longer. As before, the D2 step was held at 0.075 for all participants.

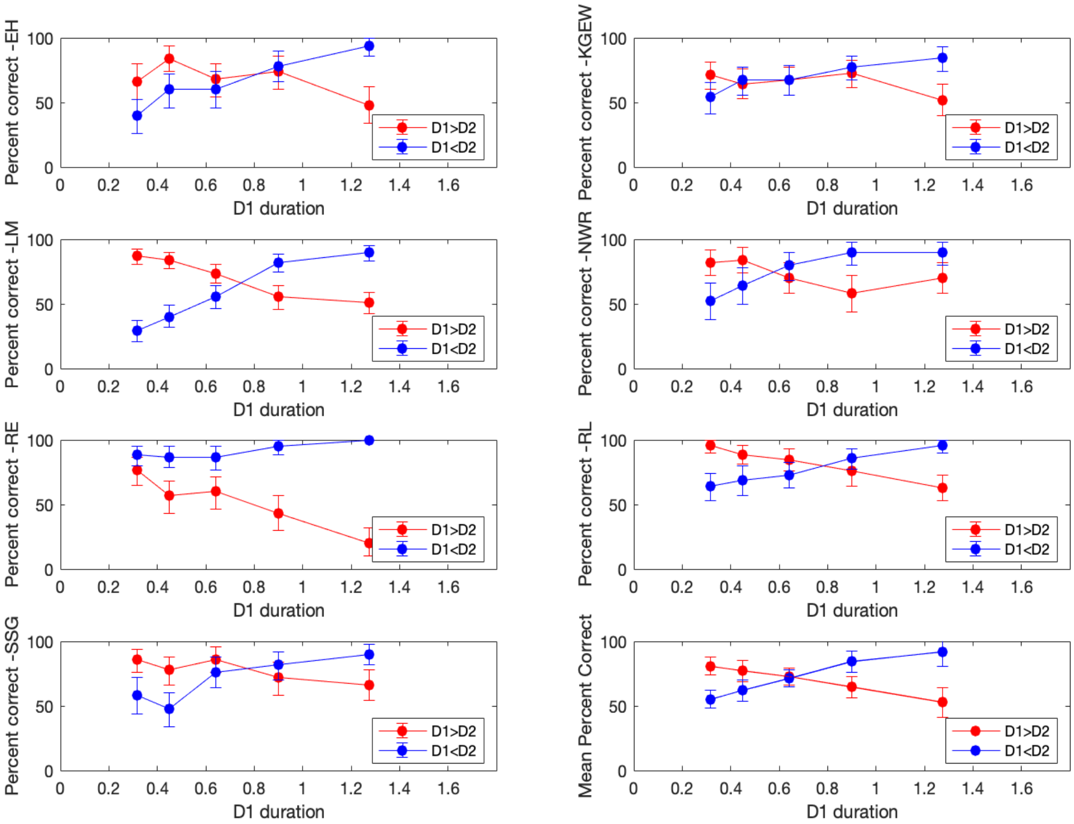
6.3 Results

During discrimination trials, participants selected the response which corresponded with the duration they perceived as longer (first or the second), and their responses were compared with the correct answer on each trial to produce a percent correct value for all trials in every run.

Figure 44 shows percent correct performance as a function of D1 for each participant, and the mean percent correct values at each duration. In each case, the red data points indicate performance when D1 > D2 and the blue data points indicate performance when D1<D2. According to time order error models which contain the mechanism of central tendency bias (or contraction bias as it is referred to in some TOE literature; Ashourian & Loewenstein, 2011; Raviv et al., 2012), participants perception of D1 will show relatively large overestimation at shorter durations. This bias would be larger than that associated with D2, because D1 carries more uncertainty due to the fading trace of D1. Therefore, when D1>D2, this bias leads to easier discriminations at small values of D1 (due to the larger overestimation bias in D1 at short durations compared to D2; see Figure 44). As D1 increases in size, D1 and D2 become more difficult to discriminate and therefore the percent correct decreases. Conversely, mechanisms of central tendency bias predict underestimation of longer durations and as before, more of this underestimation is shown in D1 relative to D2. This bias
drives the effect seen in the line plotted as D1< D2 (see Figure 44), and the trend of this line is driven by the more difficult discrimination initially of D1 and D2 when D1 is short. However, as D1 increases in size, the ease of discriminability improves and percent correct increases.

Time order error biases are clearly evident in the individual and the group averaged datasets, with percent correct increasing as a function of D1 when D1<D2 and decreasing when D1>D2. Additionally, when the data was averaged across all participants, there was a clear intersection between the two slopes occurring close to the centre of the stimulus distribution (see Figure 44).

Figure 44. The percent correct for the discrimination judgements of D1 and D2 as a function of the D1 durations for each participant and the mean percent correct calculated across all participants.

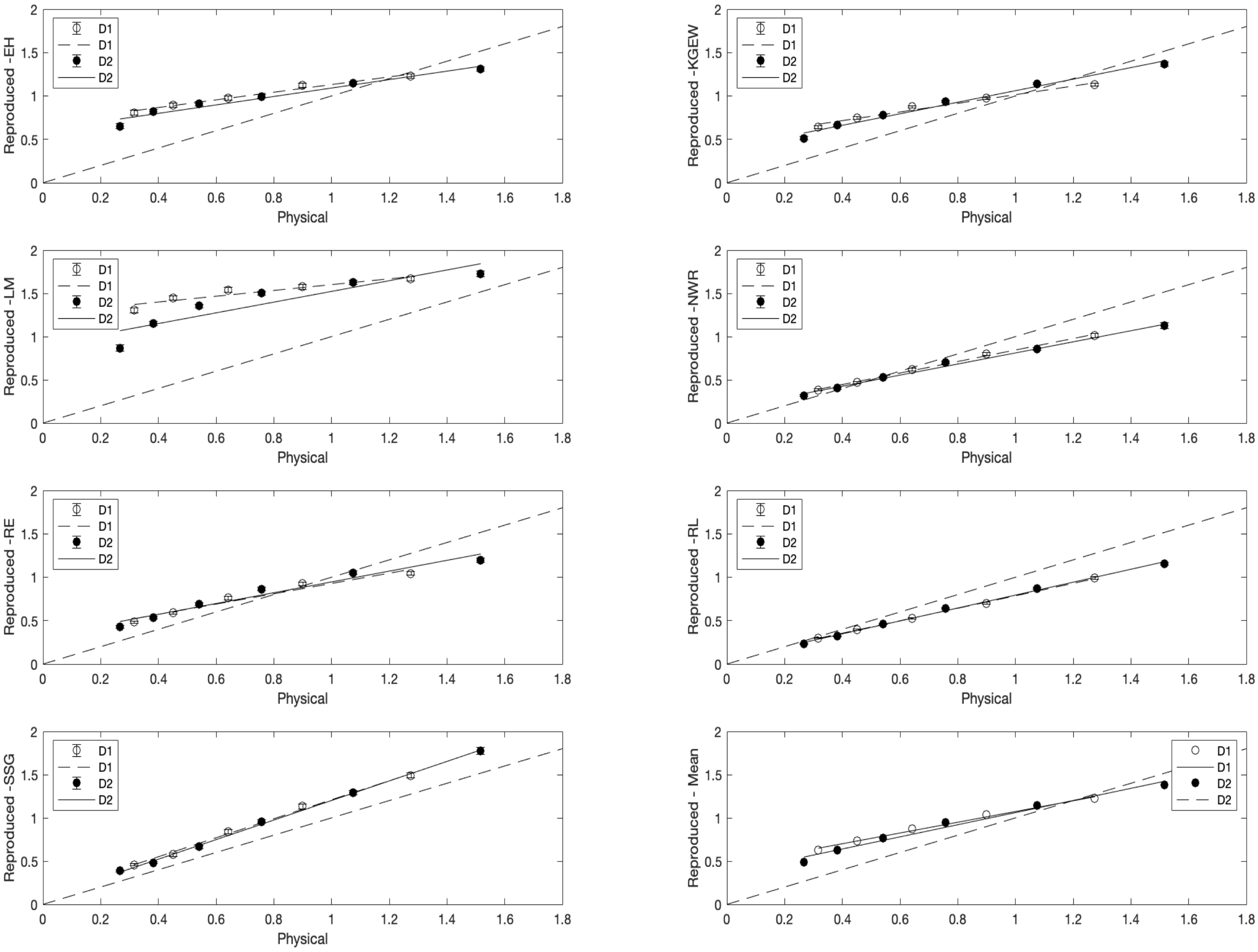


Note. Instances where D1>D2 are shown in red and instances where D1<D2 are shown in blue. Error bars represent confidence intervals.

During the subsequent reproduction trials, participants were asked to either reproduce D1 or D2. The duration they were required to reproduce depended upon which duration in the pair (D1 or D2) was replaced by a delay**.** Figure 45 shows reproduced durations of D1 and D2 plotted as a function of the respective physical stimulus duration. Linear regressions were separately fitted to D1 and D2 reproductions for each participant, with the gradient of the line allowing quantification of the degree of compressive, central tendency bias.

The individual data showed central tendency related bias in D1 and D2 for all participants to varying degrees. For the mean reproductions at each duration, bias towards the mean of the D1 and D2 stimulus distributions was also seen, with a greater tendency for overestimation of small durations than underestimation of larger durations. The strength of the bias tended to be larger in reproductions of D1 (M= 0.621) than in reproductions of D2 (M=0.701), reflected in the shallower gradient of the fitted functions. A one tailed t-test was conducted to verify this assumption made by previous research and found a significant difference t(6) = -1.979, p=0.048.

Figure 45. The reproduced durations were plotted as a function of the physical values of D1 and D2.



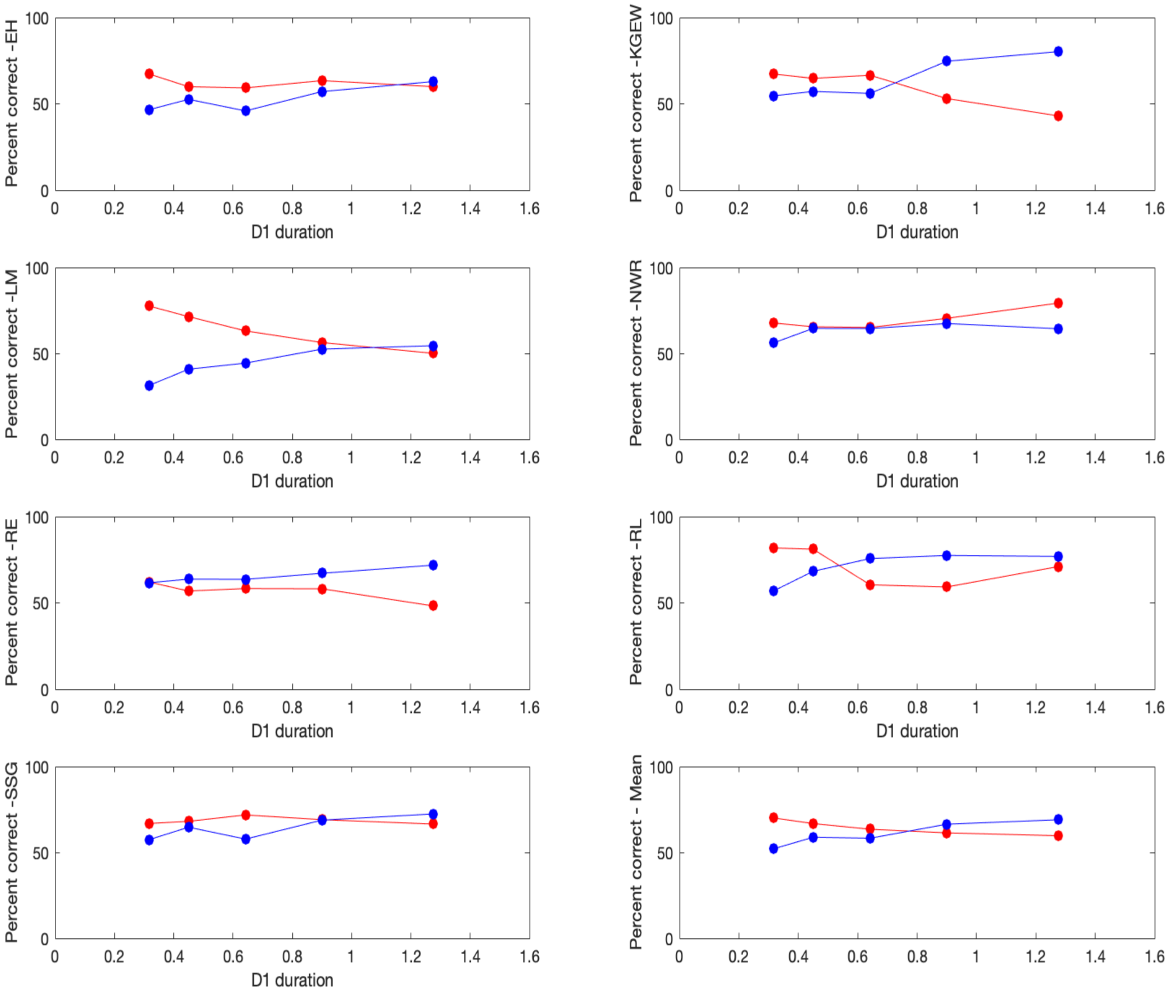
Note. A linear regression line was fitted to show the bias in the reproductions of D1 and D2

To determine whether TOEs could be directly predicted from participants' reproductions of D1 and D2, a method was needed to simulate 'discriminations' from reproduction data. First, a resampling process was used to draw a pairs of D1 and D2 reproduction trials at random from the data set (with replacement). Sampled reproductions of the D1 and D2 on each iteration were then compared to generate a simulated 'D1 longer' or 'D2 longer' response. Simulated discrimination data were than sorted into instances when the physical values of the selected stimulus pair reflected D1>D2 and instances where the values in the selected stimulus pair contained values where D1<D2. This was then transformed into a percent correct in order replicate the analysis of the pure discrimination trials. However, in this case, it was the length of reproduced durations which confirmed which duration was perceived as longer instead of a single discrimination judgment.

Figure 46 shows the simulated discrimination of the reproduced D1 and D2 durations. The data from individual participants did not provide convincing TOE bias patterns. This was further evident from group averaged data, where the two relatively flat lines were not able to demonstrate a TOE bias. An intersection of the two lines was visible but, there was little difference shown between the percent correct between D1 and D2, as D1 increased in size (see Figure 46).

The resampling process provides an intuitive method of inferring discrimination responses from reproduction data. However, it suffers from a certain limitation. Simulated discrimination is dependent upon two things 1) the difference between the mean reproduction for each and the 2) the total amount of reproduction variability. As reproduction data becomes increasingly variable, simulated discrimination performance will tend towards chance levels. While this is not problematic if variability solely reflects sensory noise, the presence of motor or other forms of noise will lead a systematic underestimation of discrimination accuracy. To avoid this, it was necessary to partition the sensory and non-sensory components of reproduction variability.

Figure 46. The reproduced durations were resampled and compared to simulate discriminations between reproduced D1s and D2s for each participant and the mean percent correct calculated across all participants.



Note. Once the reproductions provided an inferred discrimination, the percent correct accuracy was plotted as a function of the D1 duration. The mean percent correct is an average of the percent correct from all participants.

The coefficient of variation model

To account for the confounding influence of motor noise on the reproduction data, it was necessary to model the sensory and motor noise in the responses. To do this, the coefficient of variation was modelled as the total noise associated with participants reproductions. To create this model, it is assumed that sensory variability behaves according to Weber's law; that the standard deviation of duration estimates is proportional to the duration itself. Furthermore, it was assumed that motor noise was independent of sensory noise and had an additive effect. Under these assumptions, the standard deviation of reproduced durations (σ_D) is given by:

$$\sigma_D = \sqrt{\sigma_{sen}^2 + \sigma_{mot}^2}$$

where σ_{sen}^2 and σ_{mot}^2 represent the sensory and motor noise respectively.

Under Weber's law, sensory noise is proportional to the duration (D):

$$\sigma_{sen} = wD$$

Therefore, the coefficient of variation (CV) at each duration can be modelled by:

$$CV = \frac{\sqrt{(wD)^2 + \sigma_{mot}^2}}{M_D}$$

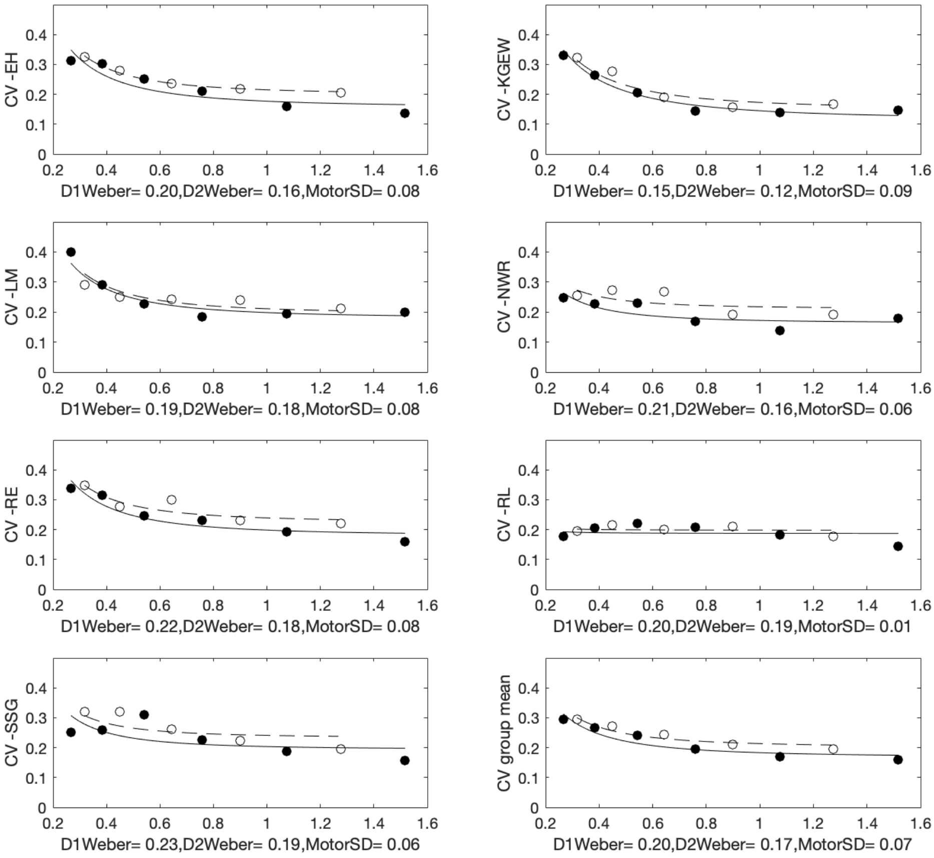
where M_D is the mean reproduction at that duration.

This model has two free parameters, the Weber fraction controlling sensory noise (w) and the motor noise (mot). The sensory noise was allowed to vary for D1 and D2 due to the predicted differences in uncertainty between D1 and D2. Motor noise was held constant across D1 and D2 due to the lack of predicted difference in motor noise across the durations. Both parameters were allowed to vary between participants to account for individual differences.

The coefficient of variations for D1 and D2 durations for each participant were plotted as a function of the duration values (see Figure 47). Most participants showed larger CVs for shorter durations and the descending slope began to plateau after 0.4 to 0.6 seconds. Under the applied model, this reflects the transition between motor noise dominating at shorter durations to sensory noise dominating at longer durations.

Once the CVs were modelled using the mean of the participants data, the variability associated with D1 reproductions appeared consistently larger than the variability associated with D2 reproductions. The CVs for the D1 reproductions were found to be significantly larger the CV for the D2 reproductions using a one tailed t-test, t(6) = 4.114, p=0.03.

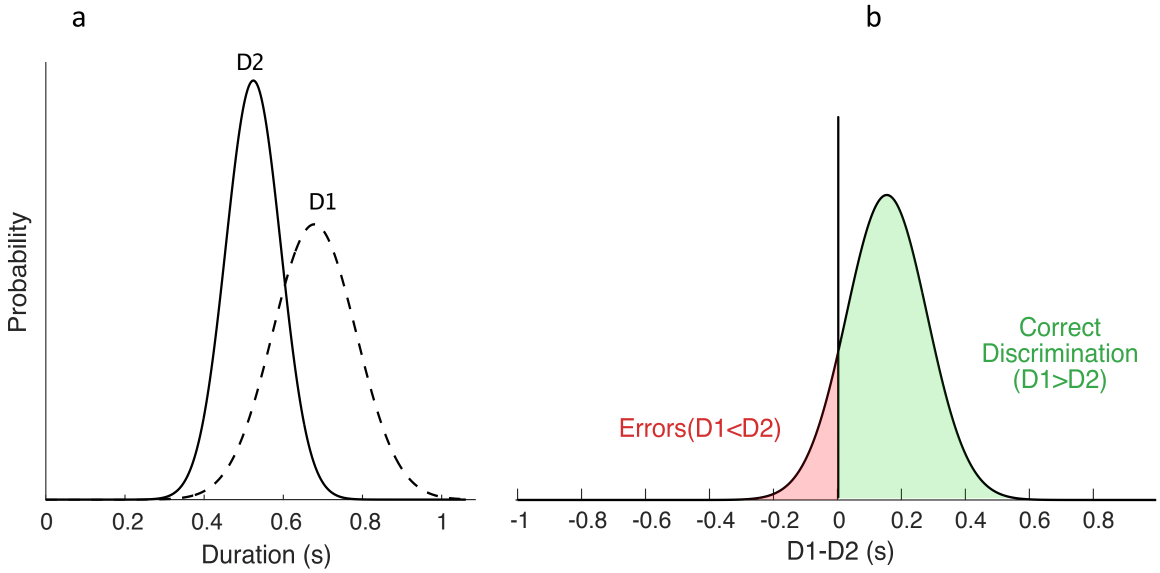
Figure 47. The coefficient of variation plotted as a function of D1 values for the reproduction procedure for each participant and the mean CV across all participants.



Note. The open circles were used for the CV associated with D1 reproductions and the filled circles were used for the CV associated with D2 reproductions. The dotted and solid lines were used for the fitted lines of D1 and D2 reproductions respectively.

By obtaining estimates of w_{D1} and w_{D2}, from the best fits of the internal estimates, the pairs of $D1$and $D2$ stimuli could be modelled. Figure 48a shows one example of where $D1 > D2$. Internal estimates were modelled as normal distributions with means corresponding to the mean reproductions (M_{D1} and M_{D2}) and standard deviations $w_{D1}D1$ and $w_{D2}D2$.

Figure 48. Duration plotted as a function of probability to show how resampling was used to infer discriminations from reproductions.



Note.

a) Resampled values of D1 and D1 may show large overlapping variability due to presence of motor noise.

b) When D1 variability is subtracted from D2 the difference is equal to the area under the curve. Any area located within the positive side of the x axis provides the probability of correct discrimination. The area under the curve that is located in the negative x axis area provides the probability of incorrect discriminations.

Given these distributions, discrimination responses can be inferred from reproduction responses by modelling the noise of the distribution (and motor correcting it) as an alternative to the resampling approach described earlier. The advantages of modelling the noise is most clearly illustrated by plotting the difference between the two distributions **(**$D1 - D2$; Figure 48b).

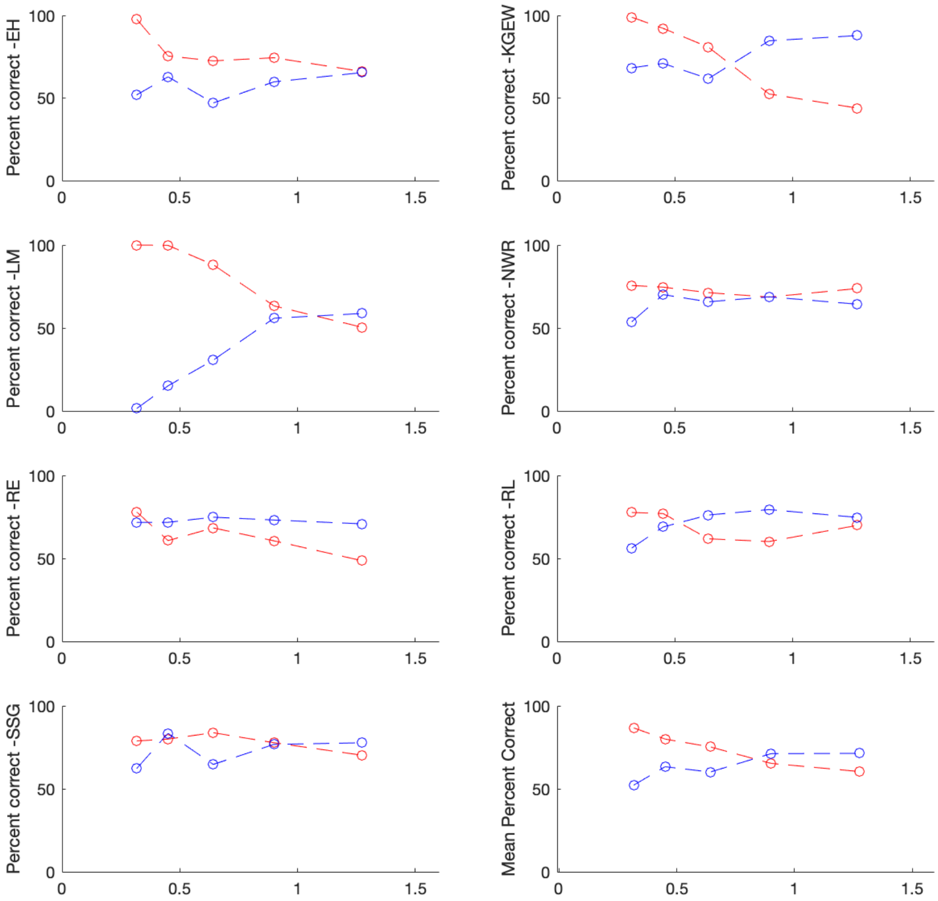
Correcting for the motor noise illuminated the remaining sensory variability (from internal representations of D1 and D2) present during a discrimination of two stimuli. This sensory noise was displayed in the sensory variability Gaussians of D1 and D2 (see Figure 48b**).** The likelihood of making mistakes (and subsequent percent correct) could be calculated based on the overlap between the sensory Gaussians for each D1 and D2. This is because when calculating the difference (subtraction) between the Gaussian curves leads to the addition of their variability squared as a mathematical rule. Therefore, the overlap between D1 and D2 becomes the area under the curve of the difference between the two curves (see Figure 48b).

Percent correct discrimination was estimated by calculating the proportion of the difference distribution falling the appropriate side of zero. When is D1 was physically larger than D2, the area on the right hand side of zero (positive) would show the percent correct. However, in cases where D1< D2 then the area of the left hand side of zero (negative) would equate to the percent correct for that discrimination of D1 and D2.

Motor-corrected reproduction data

Figure 49 shows simulated reproduction data using motor-corrected estimates of sensory noise. Results obtained using the mean percent correct data show clear evidence of a TOE. TOEs were also shown at the individual level for four participants (LM, KGEW, NWR and RL) but were not as clear for the other three participants (EH, SSG and RE). Interestingly, for RE, this lack of TOE was also reflected in their earlier discrimination data (see Figure 44).

Figure 49. Motor correction of the reproduced durations was used to provide an improved inferred discrimination for each participant and the mean percent correct calculated across all participants.



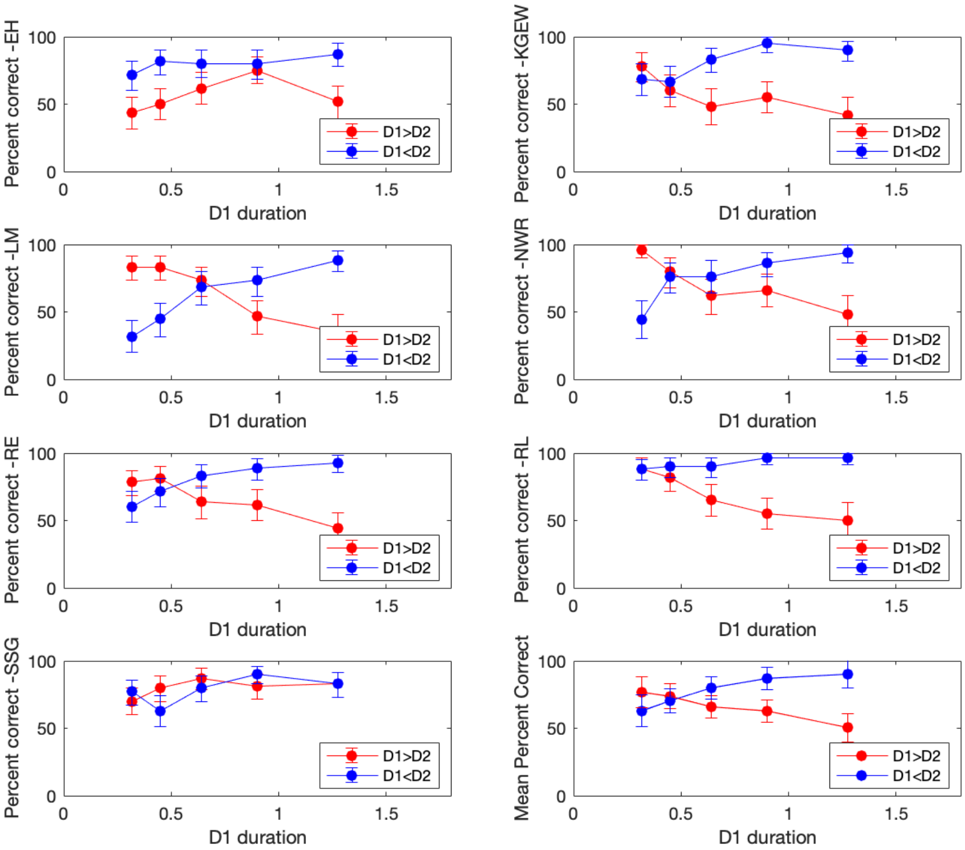
Note. These plots show the percent correct for the 'inferred discrimination' of the motor corrected reproduced durations of D1 and D2 as a function of the D1 durations. Instances where D1>D2 are shown in red and instances where D1<D2 are shown in blue.

Combined data: Discrimination

For the data described thus far, duration discriminations were inferred from reproductions of durations presented in isolation. However, there are two potential issues with comparing these simulated responses with actual discrimination judgements: (1) the discrimination and reproduction data was collected in separate sessions; and (2) stimulus sequences for reproduction judgements differed from those used for discrimination judgments in that only a single stimulus was presented. To attempt to address this, a combined method of responding was used to collect discrimination and reproduction responses simultaneously in each trial. This process involved the participants using a button press to indicate which stimulus they thought was longer and held it for a time that was equal to how long they thought it lasted.

The data were analysed using the same process as the pure discrimination trials in order to produce the percent correct accuracy for each D1 and D2 pair (see Figure 50). There is a TOE bias shown for most participants, although for 2 participants (KGEW and NWR), the intersection occurred during the shorter durations and for another two participants (SSG and RL) there was not a clearly defined intersection of the two plotted lines. For participant RL, the D1<D2 line was near 100 percent correct for the duration of the session**.** The mean percent correct data however, showed a TOE but displayed a relatively early intersection of the two lines.

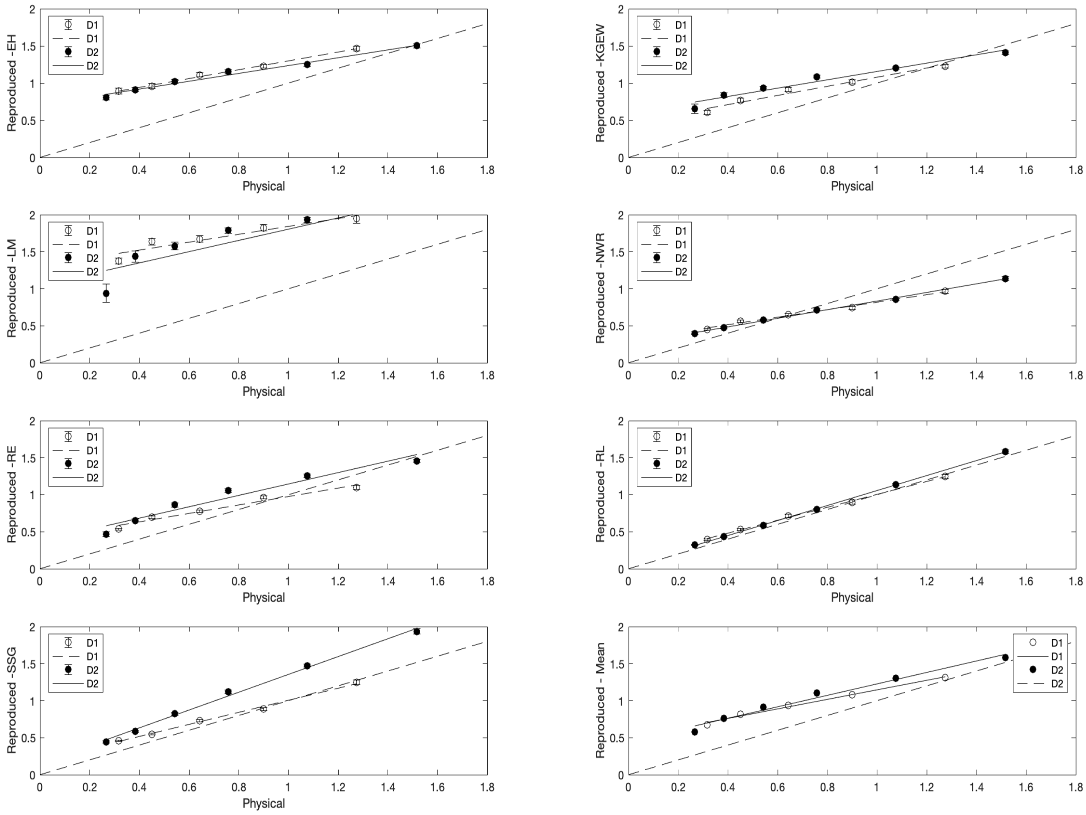
Figure 50. The percent correct plotted as a function of the D1 duration for the combined trials for each participant and the mean percent correct calculated across all participants.



Combined data: reproduction slopes

The reproduction slopes for the combined trials were produced using the same method as that used for the pure reproduction trials and were plotted for individuals and as a group average in Figure 51. Most individuals, similar to the pattern seen in the pure reproduction trials, showed more overestimation at shorter durations. Although SSG showed more overestimation at larger durations and LM continued to show large amounts of overestimation at all duration lengths for D1 and D2. For the data that was averaged across all participants, some bias towards the mean of the stimulus durations (as indicated by a decrease in the slope gradient) was visible in D1 and D2 but the gradient shown in D1 (M=0.644) was smaller (flatter slope) than in D2 (M=0.773). A one tailed t-test was significant t(6) = -2.119, p =0.039.

Figure 51. The reproduced durations of individuals and a group average, as measured during the combined procedure, plotted as a function of the physical duration values of D1 and D2.

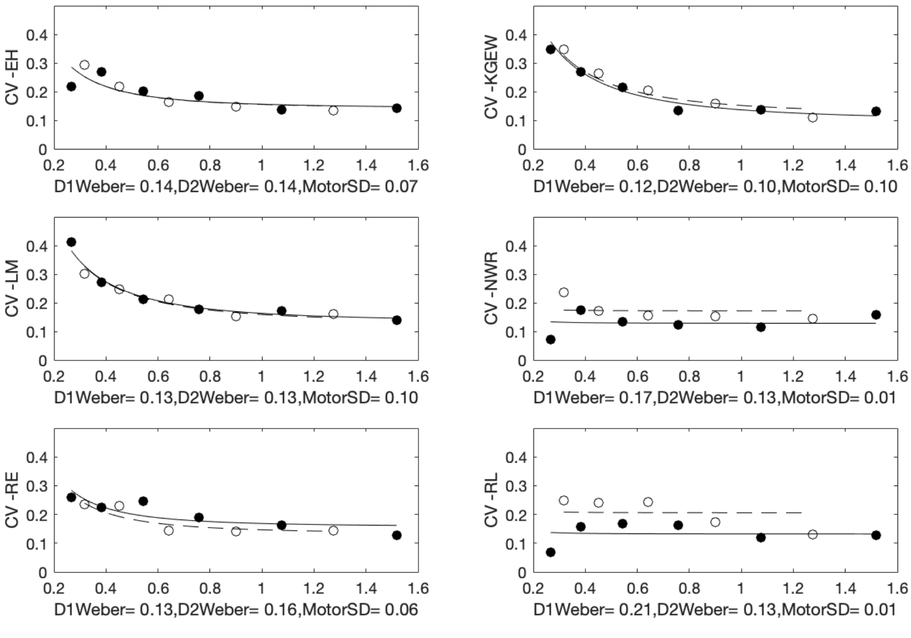


Note. A linear regression line was fitted to show the amount of bias in the reproductions of D1 and D2. In combined trials, the participants were asked to reproduce which duration they thought was longer of the durations. Therefore, they may have produced either D1 or D2 depending on their perception.

Combined data: CVs

CVs for reproductions during the combined trials were calculated as before and plotted in Figure 52. Four individual participants demonstrated a peak at short durations which later plateaued, although the other three participants showed flat CVs and therefore demonstrating a uniform relationship between the SD and M. In the group data, D1 again showed consistently higher values of CV than D2. The CVs for the D1 reproductions was not found to be significantly larger the CV for the D2 reproductions using a one tailed t-test, t(6) = 1.144, p=0.148.

Figure 52. The coefficient of variation plotted as a function of D1 values for the combined procedure for individual participants and the group mean.



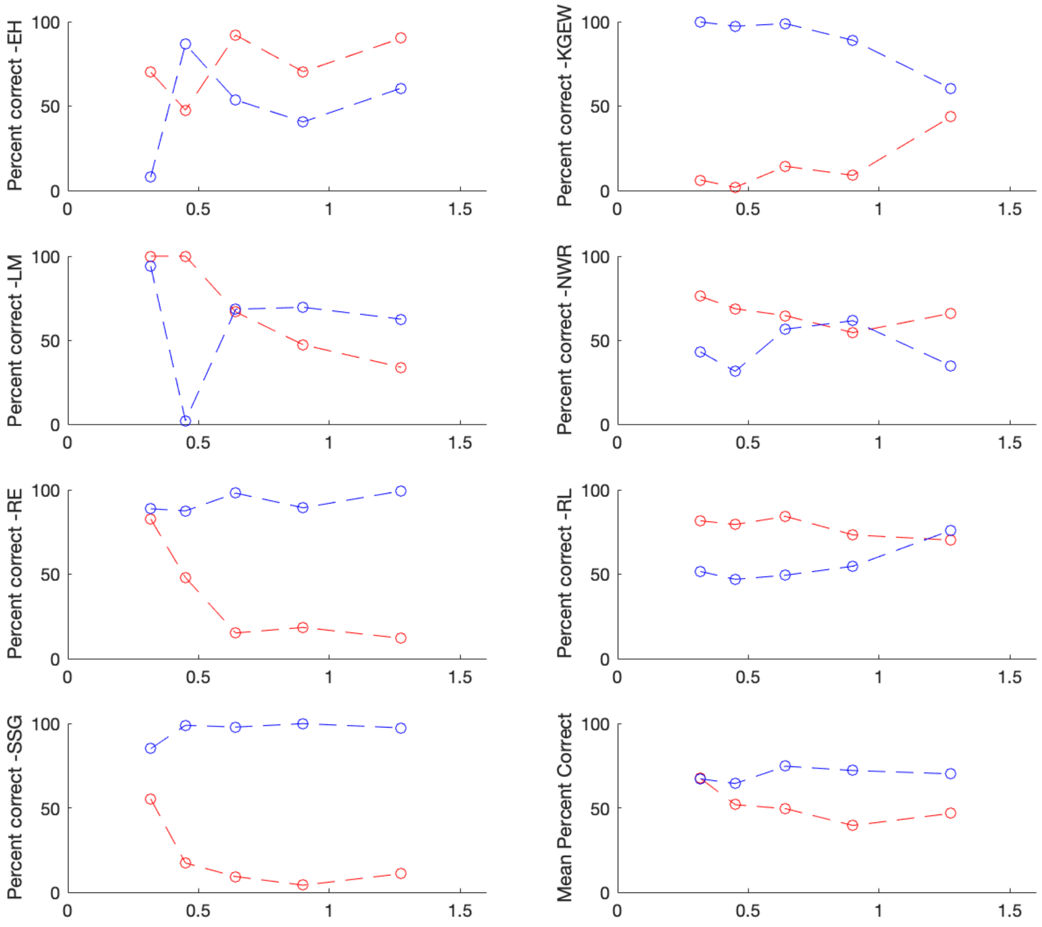
Note. The open circles were used for the CV associated with D1 reproductions and the filled circles were used for the CV associated with D2 reproductions. The dotted and solid lines were used for the fitted lines of D1 and D2 reproductions respectively.

Combined data: reproduction motor corrected.

The data for the motor corrected reproductions were analysed using the CV model that was used for the pure reproduction trials and were plotted in Figure 53. The individual data and the 'mean percent correct' figure, the data did not display a clear TOE. The intersection of the two lines was only present at the first value of the stimulus distribution. However, this was also true of the discrimination data TOE that was analysed from the combined trials (see Figure 50).

The individual data patterns reflected the lack of TOE seen in motor corrected reproduction for 2 participants because the two participants with a weaker TOE showed less over and underestimation of D1 (SSG and RL; see Figure 51). Although this was not true for KGEW and RE who still showed visible under and over estimation of D1 (see Figure 51), but no visible TOE effect in the percent correct data (see Figure 53).

Figure 53. Motor correction of the reproduced durations was used to provide an improved inferred discrimination.



Note. The percent correct of the inferred discrimination data plotted as a function of the D1 duration for the combined trials procedure for each participant and the mean percent correct calculated across all participants. The instances where the physical values of D1>D2 were plotted in red and the instances of D1<D2 were plotted in blue.

6.4 Discussion

The current experiment aimed to test the ability of a Bayesian framework to account for time order errors in binary discrimination judgements. More specifically the ability of Bayesian model to explain how memory degradation, which leads to a larger contraction bias of D1 than D2, eventually culminates in a TOE. Logically, discrimination of a two stimulus comparison should be subject to the same memory degradation processes as reproduction of a single stimulus (where either the first or the second stimulus of the pair was replaced with a delay). At this current point in time, the prevailing SW model does not have an explanation for this. Discrimination and reproduction data were directly compared to investigate this possibility. From the discrimination data, participants showed a time order error that is, the error was driven by an increased central tendency bias for the perception of D1 compared to D2. The production of a time order error was verified by the observation that discriminating instances where D1<D2 became more difficult as D1 increased but D1>D2 became easier as D1 increased. Resulting in the expected cross shape of the two plotted lines.

The production of a TOE in the current research supported previous literature. Previous empirical data has historically been interpreted through the sensation weighting model (SW). The SW model described that participants compared a weighted version of the two stimuli and not the exact stimulus values (Hellström, 1979, 2000; Laurie Anne Sapey-Triomphe, Timmermans, et al., 2021). The uneven assignment of the weights to each stimulus accounted for the findings that the first interval is relatively more biased (over or underestimated) in the perception of its size than the second. The fading trace account suggested this was due to how the memory of the first interval has degraded relative to the second. This uncertainty associated with the first stimulus has since been connected to Bayesian explanations (Ashourian & Loewenstein, 2011; Hellström & Rammsayer, 2015; Raviv et al., 2012). As a result, this research aimed to understand whether the same fading trace account, and subsequent Bayesian interpretations of fading trace (Ashourian & Loewenstein, 2011; Raviv et al., 2012), could provide an explanation of TOE during reproduction.

Participants' reproductions clearly displayed central tendency bias in both D1 and D2. However, the amount of bias for D1 was larger than the amount of bias in D2 as displayed in the group data. This was in line with fading trace accounts of time order error but, does not directly reveal any difference in the ability of sensation weighting or Bayesian inference models to account for TOE bias. Additionally, the coefficient of variation (CV) of participant reproductions supported the 'fading trace' of D1 because there was more variability associated with reproductions of D1 compared to D2. This larger variability associated with D1 reproductions was shown to be significant in purely reproduction trials but was not significant for combined trials. Differences between the reproduction and combined trials are outlined later in this discussion.

The CVs were also used to model the sensory and motor noise in participants reproductions. This modelling process facilitated the ability to infer discrimination performance from the motor corrected reproduction data. The amount of bias seen in D1 and D2 was reflected during inferred discrimination by the presence of a clear TOE effect. Therefore, single reproductions of D1 revealed larger bias than reproductions of D2 and this could be driven by the fading trace mechanism (as it already exists within the SW model). This increased bias associated with D1 was significant in a one tailed paired samples t-test. Regardless, the sensation weighting model, and its focus on explaining the TOE through comparison, means that it does not provide an explanation for the presence of a TOE being produced during the reproduction of single stimuli.

Bayesian modelling research has predominantly focused on the reproduction of single stimuli (Glasauer & Shi, 2021b; Jazayeri & Shadlen, 2010) and has not often provided an explanation of how two successive stimuli may be perceived. Through a Bayesian explanation of TOE however, the first stimulus contains relatively more sensory uncertainty as the memory trace fades. The ability of the Bayesian framework to explain time order error bias from this reproduction experiment means that it might account for TOE more comprehensively than the sensation weighting model.

Resampling was the initial approach to analysing the reproduced durations in a form that was analogous to the analysis of the discrimination trials. This approach did not reveal a time order error in the group resampled data. However, the presence of additional motor noise in reproductions that was not present during discriminations. This made it difficult for the resampled durations to display true differences between D1 and D2 reproductions because they were obscured with additional noise. Therefore, the coefficient of variation was used to provide a model of total variability and the contribution of motor noise could be corrected for. The motor corrected reproductions (and subsequent inferred discriminations) provided the expected TOE effect. This demonstrated the necessity of future research in the area of TOE reproductions to consider the impact of motor noise. Overall, TOE may be present in reproduction tasks, in addition to discrimination tasks, and the Bayesian models shows promise in being able to provide a mechanism for these processes.

The combined trials procedure demonstrated that the TOE effect was not visible when the task incorporated discrimination and reproduction within the same trial. If this were also the case for the reproduction only trials, then the comparison of two stimuli would have been necessary to produce a TOE (and some explanation making reference to interference or comparison). However, the specific requirement to reproduce and discriminate simultaneously (and the associated cognitive load) led to the absence of the TOE. This suggested that passive memory mechanism of the fading trace account may not provide a sufficient explanation. As a result, future research should explore cognitive explanations relating to D1 and D2 interference and its potential necessity in producing the effect.

Additionally, in the combined trials, participants were explicitly asked to reproduce only the ‘longer’ duration. For time order error bias, D1 more frequently appears longer than D2 at the shorter durations used here (positive TOE; Allan, 1977). This could mean that when asked which duration was longer, there were fewer responses of D1<D2 relative to the responses for D1>D2. Fewer observations will have produced noisier estimates and so the added noise worked to reduce any differences between the two plotted lines (D1>D2 and D1<D2). Finally, any differences between the reproduction and the combined trials could be produced by a change in task difficulty, for example ceiling effects being present if combined trials were too difficult.

Future research could investigate the importance of interference by directly comparing any difference between reproductions of a single duration under circumstances when 1) only a single duration and a delay are present and 2) when both durations are presented but participants are only asked to reproduce a single one each time (either D1 or D2 as decided prior to the start of the experiment and not the interval perceived as longest on each trial as was the case with our combined trials). Given the support provided here for the importance of the fading trace of the first duration, it would be useful to continue to understand the impact of the ISI on the size of the TOE and whether interference becomes a larger factor in the TOE depending on the size of the ISI.

In sum, the time order error bias may be driven by a mechanism that is not exclusive to comparative estimations, as it was visible during reproduction trials. The Bayesian model seemed a likely candidate to explain this error due to its ability to differentially weight the different stimuli according to their uncertainty. This model was successful because the increased uncertainty (due to memory degradation) associated with D1 relative to D2 was seen to drive the TOEs during discrimination and reproduction. This means that the increased bias driving the TOE could be accounted for by a Bayesian inference model in the current research because the first stimulus was more susceptible to central tendency bias than the second stimulus irrespective of whether it was compared or reproduced. Although the specifics of these mechanisms remain unclear as combined trials did not produce the expected TOE bias).

7. General discussion.

7.1 Summary

The merits of the Bayesian observer model in accounting for magnitude estimations have been demonstrated throughout this thesis. Additionally, the potential for prior acquisition strategies to vary has been explored, in relation to the characteristics of the stimulus sequence, different stages of development and individual differences in autistic traits. Previous research that incorporated Bayesian prior learning informed magnitude estimation biases such as multisensory cue combination (Körding et al., 2007), visual motion (Alan A. Stocker & Simoncelli, 2008), path integration (Petzschner & Glasauer, 2011) and timing behaviour (Jazayeri & Shadlen, 2010). Relatively speaking, less is known about exactly how priors develop, but an iterative model has demonstrated the best fit to participants data in recent literature (Glasauer & Shi, 2021b; Petzschner & Glasauer, 2011). Overall, this thesis demonstrated that different prior acquisition strategies may be suited to different stimulus features and operate differently across temporal and spatial domains. Additionally, some evidence has been found for variation of strategy between adults and children, but to a lesser extent neurotypical participants with differing levels of autistic traits. These conclusions will be elaborated on below in relation to several sub questions.

7.2 How are Bayesian priors learned in the general population?

Bayesian prior weighting has provided a successful explanatory mechanism for the central tendency bias shown throughout this thesis. Findings pertaining to how priors were acquired and formed over time seemed to conflict with recent research. In Glasauer and Shi (2021) a fixed Kalman filter strategy was shown to be the most likely strategy for temporal estimation because there was a significant reduction in central tendency bias for autocorrelated stimuli relative to fully random stimuli. However, In Chapter 2 of this thesis, a reduction in compression for autocorrelated duration stimuli was not shown. Additionally, the model that provided the best fit for both unfilled and filled durations was the static model. Therefore, it is possible that more than one prior acquisition strategy for time perception exists, but the static prior provided the best explanation for the data set shown in Chapter 2. Future research will be needed to determine the timeframe for a switch in strategy. Under conditions where participants are not given enough time, a switch may not be seen.

The lack of support for an iterative model in Chapter 2 may have occurred due to certain cues in the experimental environment. For example, in Glasauer and Shi (2021b)'s experiment they might have provided certain conditions (such as a variable step random walk) that pushed participants towards a tipping point, where the static model no longer provided the best approach and, so participants deviated away from a static model and towards a Kalman filter. The concept of a 'tipping point' is strengthened by the relative success of the two state model compared to the Kalman filter in the current research. The small advantage provided by the two state model relative to the static model, due to its ability to capture a shifting prior mean. Although it is important to note that for the current research, this ability of the two state model to explain the data was very similar to the static model. This suggested that the tipping point into a fully iterative model was not met. Additionally, the unexpected findings in relation to object type, also called into question the assumption that temporal characteristics of single objects were viewed as more autocorrelated than a group of objects. Here, it was assumed that the spatial characteristics of objects (whether multiple or single objects were presented) were related to the perceived temporal characteristics. The object type effect in Chapter 2 was seen in the opposite direction to what was expected so it was possible that either 1) participants did not pick up on the single and multiple object cues or 2) that the spatial object characteristics were not viewed as relevant to their temporal characteristics. In future research it may be useful to inform participant about the nature of the objects as to which ones are to be treated as single or multiple objects.

Furthermore, there was no clear relationship between bias and variability in the temporal reproduction data. The compression magnitude bias showed differences across conditions, but individual participant variability was highly similar across the different conditions. A relationship between bias and variability was expected because both values are impacted by changes in sensory noise under Bayesian integration. Also, in simulations of iterative models, even if sensory noise is unchanged, the stimulus sequence (degree of autocorrelation) impacts bias and variability. It is possible then that a different unknown factor may have led to increased weighting of the prior (not the participant variability) in certain conditions. Or alternatively, that participants were not operating in a way that was optimal under the Bayesian framework. In sum, the data from Chapter 2 provided insufficient evidence to claim the operation of a Kalman filter, as a static model provided the best fit for most participants. Overall, this could mean there is more than one possible strategy for temporal prior learning, given the lack of agreement with previous findings.

To investigate the possibility that a ‘previous target’ prior learning strategy may be better suited to spatial estimations, a series of spatial localisation experiments were conducted in Chapter 4. When fitting the models simultaneously to all conditions, the previous target model provided the best fit. However, when fitting the models separately to the two different stimulus sequences, the best fitting prior updating strategy appeared to be flexible according to the stimulus features present. The previous target model was most successful when the models were fitted to the random walk sequenced stimuli, but the moving average model then became the best fitting model for the fully random sequences. The observed ‘switch’ in the best fitting model could have occurred because previous target strategies may have provided a more successful way to predict the next value for autocorrelated sequences. This is the first demonstration that prior acquisition strategies could change to match the stimulus features or sensory input.

As described earlier, a static model may be the default for timing because the iterative model showed a very poor fit in Chapter 2 and despite the relatively good fit of the two-state model, it provided little improvement upon the static model for all participants. Demonstrating that the iterative element of the two state model added very little to the prior acquisition process. However, spatial quantities seen in natural settings may frequently display autocorrelation because spatial position of objects tends to vary gradually as opposed to categorically or sporadically. As a result, the appropriate model for spatial estimations might favour the ability to track gradual change. This could be related to the difference in the natural statistics of spatial and temporal inputs. It can be deduced that in the spatial domain, the locations of an object over time as it moves is likely to be autocorrelated and so too are associated features such as speed. However, this is not the case for temporal quantities and so there is less reason for participants to view temporal duration in this manner. It is possible that, compared to temporal magnitudes, this could tip the scales of a 'default' general static strategy because there were enough cues in the spatial localisation stimuli to suggest a previous target strategy was needed.

A combination of both strategies may also provide a plausible method for prior learning during spatial judgements. The two state model used a combination of static and iterative strategies to provide a relatively good fit to temporal estimation data. Although the transition parameter, which was designed to reflect a trial to trial shift in the mean, was close to 0 for most participants. Therefore, in the current research, a static model provided a relatively strong candidate for the participants strategy. Although the strength of a combined strategy (two state model) was not explored in relation to the spatial data because the necessary changes to the model were beyond the scope of the current thesis given the large differences in how uncertainty was presented between the two tasks. Future research would be needed to explore strategies that exist in the space between iterative and static strategies. This could involve exploring the influence of previous trials (using an N-back analysis as described in Hallez et al., 2019) in more detail and, potentially contrasting this with use of the mean of the previous 2 to 5 trials to form priors, in a way that was similar to the design of Lieder and colleagues (2019). The final difference between the timing and spatial task was the feedback provided during each trial of the spatial task. This may have influenced participants ability to switch prior updating strategies, when deemed necessary, in line with said feedback.

7.3 How may prior learning vary i) throughout development and ii) with autistic traits?

i) Prior learning throughout development

The aim of Chapter 5 was to explore differences in spatial localisation between adults and children aged 6 to 11 years old. The amount of variability shown in participant responses decreased with age because both groups of children showing significantly larger variability than the adults. The bias towards the mean of the stimulus distribution, also decreased with age, with the youngest children (6-8 years) showing significantly larger bias than the older children (9-11 years old) and the adults. Crucially, the amount of bias was not significantly different between older children and adults. This demonstrated that the amount of prior weighting displayed during this task reaches an adult level between the ages of 9 and 11. The concurrent decrease in the bias and variability shown here supported the findings of Karaminis and colleages 2016, although in the current research, bias decreased more quickly than variability during development.

The previous target model was favoured during random walk sequences for participants of all age groups when the data was fitted simultaneously. Therefore, the previous target initially appeared to be the default model across conditions and ages during spatial localisation. Once the models were fitted separately to each condition, the model fits shown supported the idea that adults could be switching their model according to the stimulus sequence cues. However, only some of the adults displayed this switching pattern. Chapter 4 aimed to reveal whether children may show more, or less, flexibility than adults. For older children previous target model provided the best fit in all cases showing them to be relatively more rigid than the adults. However, the younger children showed some flexibility, but this was likely a reflexive response to the type of stimulus sequence they were presented with first. The adult group were presented with more trials and were tested in their home environment through Pavlovia. Therefore, the number of trials and features of the testing environment may be relevant for future research to explore in relation to flexibility.

ii) Prior learning and autistic trait scores.

In Chapter 3, AQ trait scores were correlated with duration reproduction data. In terms of the relationship between AQ scores and compression magnitude, the random walk conditions provided the only significant correlation. This correlation was weakly negative but may suggest that those with high trait scores, might have reduced their compression when presented with autocorrelated stimuli to a greater extent than those with lower trait scores.

The correlation between participant variability and autistic trait scores was not significant. If as, Brock (2012) suggested that autistic individuals may display more precise sensory likelihoods, this would be reflected in participant variability. Therefore, it is unlikely that the higher trait individuals here had more precise likelihoods because they showed more noise in their responses than lower trait individuals. Increased noise, more specifically motor noise, in high trait individuals could be related to the increased levels of motor difficulties seen in actually autistic individuals (Green et al., 2009; Licari et al., 2020; Wang, Petrulla, Zampella, Waller, & Schultz, 2022).

A negative correlation between Akaike weights and high AQ traits may have been expected, given how the reduced compression found for random walk sequences in high AQ individuals implied Kalman filter use. However, upon further exploration of model fits on a trial by trial basis, this was not the case. Akaike weights provided a metric for how well the models were fitted to the data for both static and iterative models, where closer to 1 indicated a static model. The relationship between the quality of fit of the Kalman filter, as determined by the Akaike weight, and AQ score did not provide a significant correlation. Taking into consideration that most participants in the sample were favouring the static model, then showing an increased use of an iterative model for higher trait individuals would have been unlikely. Furthermore, due to the small number of high AQ individuals this may not have been a true reflection of autistic perceptual experience so future research in this area would benefit from selecting high trait and or diagnosed autistic individuals. The importance of this distinction has been shown in recent research because a correlation of AQ traits with central tendency bias was not found but significant differences were shown for those diagnosed with autism compared with non-autistic participants (Theisinger, 2022).

7.4 Comparing the mechanisms underlying biases during discrimination and reproduction.

More recent Bayesian theories of autistic perception have proposed that priors may have been formed but differences may exist in the ability to flexibly adjust the prior precision. Sapey Triomphe and colleagues (2021) demonstrated priors of autistic individuals were strong and less flexibly adjusting to the context using of the Time Order Error (TOE). This is because they did not update their prior precision to account for a wider stimulus range. Therefore, TOE was used by the authors to show how Bayesian prior distribution learning was less flexible for autistic individuals.

Sapey Triomphe and colleagues (2021) assumed that memory degradation led increased bias and weighting of the first stimulus to produce the TOE bias. In the current thesis, the logic of this assumption was tested in relation to the reproduction of single stimuli in Chapter 6. This was deemed necessary as an interim step prior to exploring the Bayesian prior formation process in autistic individuals because the assumptions of the TOE itself needed to be investigated before continued use this task for understanding autistic prior learning. The core assumption of the Bayesian model is that there is larger uncertainty (and therefore larger bias) associated with the first interval during the comparison of two weighted stimuli. The current prevailing model of TOE, sensation weighting, and the Bayesian model both implement a mechanism of differential weighting of the first and second stimulus. However only the Bayesian model allowed this assumption to be tested in relation to reproduction of the stimulus duration.

In Chapter 6, reproduction bias was greater for the first stimulus than the second stimulus and a TOE was shown using inferred discriminations from a motor noise corrected distribution of participant reproductions. This finding successfully demonstrated the relevance of stimulus bias outside of a two stimulus comparison process. However, it was necessary to understand the relevance of two factors: 1)both stimuli appearing in the same trial and 2) discrimination and reproduction occurring during the same experimental session. During these combined trials the time order effect was not present, which could be related to the presence of both discrimination and reproduction during the same trial. One reason for this could be that combined trials were more difficult than other trial types.

Overall, findings from the reproduction only trials of Chapter 6 suggested that the presence of two intervals was not necessary for the TOE effect (and associated biases) to be produced. This provided strength for a Bayesian interpretation of the increased uncertainty associated with reproducing the first interval and the resultant increased bias in the first stimulus relative to the second. The continued use of a Bayesian framework may assist in understanding the TOE bias in future research.

7.5 Conclusion.

An overarching aim of this thesis was to bring together Bayesian observer models and the concepts learned from developmental psychology. Until more Pellicano and Burr (2012)'s suggestion, these areas seemed to be studied separately and based on their suggestions and further papers, it became increasingly clear how necessary it is to unite thinking in the literature. The current thesis brought together these areas through studying perception using the Bayesian observer models in typical populations (Chapter 2 and Chapter 4) and according to AQ traits (Chapter 3) and different ages (Chapter 5).

In relation to the temporal findings, the static model was favoured for temporal duration estimation in the current research, which was in contrast to the findings of recent literature (Glasauer & Shi, 2021b). This could reflect a default static strategy for temporal reproduction and the existence of a 'tipping point' towards an iterative strategy when faced with certain stimulus cues. It could be that the more explicit cues of autocorrelation that exist in natural spatial objects (such as location and speed) could make a switch to a previous target model more likely. Or in a similar vein, it could be that a variable step random walk (used by Glasauer and Shi (2019; 2021b) and the spatial localisation task in Chapter 4), provided the appearance of stimuli that changed in a more natural and gradual way than the relatively more discrete fixed step stimuli.

In contrast, the spatial localisation experiments in Chapter 4 showed that adult participants favoured a previous target model. Although, the best fitting model switched to a moving average model when stimulus sequence changed. A switch in strategy, such as this one, during the course of the experiment has not been demonstrated in previous literature. The ability of the current research to capture developmental differences was limited in scope to a neurotypical population. Although in the developmental trajectory data set, it was shown that older children were not able to switch strategy and younger children favoured which ever stimulus cues that were contained within the first task they completed. Therefore, any real ability to switch was not present until adulthood. During adulthood, the strategy switch occurred for some participants but not all.

Another opportunity for the exploration of developmental differences was provided in the form of autistic trait correlations. Reduced compression bias was shown for autocorrelated stimuli in high trait participants. However, no other relationships were found and so the inclusion of high trait (AQ score of 32 or more) or autistic participants is crucial for future research. Overall, Bayesian prior models provided a good fit for neurotypical adults and demonstrated how several different prior acquisition strategies may be learned in response to different stimulus features. However, the ability of these models to capture trends within different neurodevelopmental conditions or trajectories remains to be seen.

References

Abel, S. M. (1972a). Discrimination of Temporal Gaps. *The Journal of the Acoustical Society of America*, *52*(2B), 519–524. https://doi.org/10.1121/1.1913139

Abel, S. M. (1972b). Duration Discrimination of Noise and Tone Bursts. *Journal of the Acoustical Society of America*, *51*(4B), 1219–1223. https://doi.org/10.1121/1.1912963

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience*, *7*(10), 1057–1058. https://doi.org/10.1038/nn1312

Adams, W. J., Kerrigan, I. S., & Graf, E. W. (2010). Efficient visual recalibration from either visual or haptic feedback: The importance of being wrong. *Journal of Neuroscience*, *30*(44), 14745–14749. https://doi.org/10.1523/JNEUROSCI.2749-10.2010

Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Allan, L. G. (1977). The time-order error in judgments of duration. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *31*(1), 24–31. https://doi.org/10.1037/h0081647

Allman, M., & Falter, C. (2015). Abnormal timing and time perception in autism spectrum disorder? A review of the evidence. In *Time distortions in mind–temporal processing in clinical populations* (Vol. 15).

Allman, M. J. (2011). Deficits in temporal processing associated with autistic disorder. *Frontiers in Integrative Neuroscience*, *5*(March), 1–2. https://doi.org/10.3389/fnint.2011.00002

Allman, M. J., Pelphrey, K. A., & Meck, W. H. (2012). Developmental neuroscience of time and number: Implications for autism and other neurodevelopmental disabilities. *Frontiers in Integrative Neuroscience*, *6*(FEBRUARY), 1–24. https://doi.org/10.3389/fnint.2012.00007

Allman, M. J., Penney, T. B., & Meck, W. H. (2016). A brief history of "The Psychology of Time Perception." *Timing and Time Perception*, *4*(3), 299–314. https://doi.org/10.1163/22134468-00002071

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The “oblique effect” in man and animals. *Psychological Bulletin*, *78*(4), 266–278. https://doi.org/10.1037/h0033117

Arnold, D. H., Tear, M., Schindel, R., & Roseboom, W. (2010). Audio-visual speech cue combination. *PLoS ONE*, *5*(4). https://doi.org/10.1371/journal.pone.0010217

Ashourian, P., & Loewenstein, Y. (2011). Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0019551

Baron-Cohen, S., O’Riordan, M., Plaisted, K., & Driver, J. (2001). Superior Visual Search in Autism. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 27, pp. 719–730.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Bausenhart, K. M., Bratzke, D., & Ulrich, R. (2016). Formation and representation of temporal reference information. *Current Opinion in Behavioral Sciences*, *8*, 46–52. https://doi.org/10.1016/j.cobeha.2016.01.007

Behar, I., & Bevan, W. (1961). The Perceived Duration of Auditory and Visual Intervals : Cross-Modal Comparison and Interaction. *The American Journal of Psychology*, *74*(1), 17–26.

Bejjanki, Vikranth R., Knill, D. C., & Aslin, R. N. (2016). Learning and inference using complex generative models in a spatial localization task. *Journal of Vision*, *16*(5), 1–13. https://doi.org/10.1167/16.5.9

Bejjanki, Vikranth Rao, Clayards, M., Knill, D. C., & Aslin, R. N. (2011). Cue integration in categorical tasks: Insights from audio-visual speech perception. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0019812

Bellville, S., & Mottron, L. (1993). A study of perceptual analysis in a high-level autistic subject with exceptional graphic abilities. *Brain and Cognition*, Vol. 23, pp. 279–309.

Berniker, M., & Kording, K. (2009). Bayesian Models of Motor Control. *Encyclopedia of Neuroscience*, *2*, 127–133. https://doi.org/10.1016/B978-008045046-9.01429-

7

Berniker, Max, Voss, M., & Kording, K. (2010). Learning priors for bayesian computations in the nervous system. *PLoS ONE*, *5*(9), 1–9. https://doi.org/10.1371/journal.pone.0012686

Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., & Stone, W. L. (2003). Visual recognition of biological motion is impaired in children with autism. *Psychological Science*, *14*(2), 151–157. https://doi.org/10.1111/1467-9280.01434

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, *134*(3), 330–343. https://doi.org/10.1016/j.actpsy.2010.03.006

Brock, J. (2012). Alternative Bayesian accounts of autistic perception: comment on Pellicano and Burr. *Trends in Cognitive Sciences*, *16*(12), 573. https://doi.org/10.4327/jsnfs1949.11.5

Buhusi, C. V., & Meck, W. H. (2002). Differential effects of methamphetamine and haloperidol on the control of an internal clock. *Behavioral Neuroscience*, *116*(2), 291–297. https://doi.org/10.1037/0735-7044.116.2.291

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, *198*(1), 49–57. https://doi.org/10.1007/s00221-009-1933-z

Burr, D. C., & Morrone, C. (2006). Time Perception: Space-Time In The Brain. *Current Biology*, *16*(5), 171–173. https://doi.org/10.1016/j.cub.2006.02.042

Caramelli, N., Setti, A. M., & Donatella, D. (2004). *Concrete and abstract concepts in school age children*. *8*(2), 19–34.

Chambers, C., Sokhey, T., Gaebler-Spira, D., & Kording, K. P. (2018). The development of Bayesian integration in sensorimotor estimation. *Journal of Vision*, *18*(12), 1–16. https://doi.org/10.1167/18.12.8

Chen, L., & Vroomen, J. (2013). *Intersensory binding across space and time : A tutorial review*. (May), 790–811. https://doi.org/10.3758/s13414-013-0475-4

Chrysaitis, N. A., Jardri, R., Denève, S., & Seriès, P. (2021). No increased circular inference in adults with high levels of autistic traits or autism. *PLoS Computational Biology*, *17*(9), 1–17. https://doi.org/10.1371/journal.pcbi.1009006

Church, Russell, M. (1984). *CHURCH- 1984 - Annals of the New York Academy of Sciences - Wil.pdf*.

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 of the United States of America*, *111*(21), 7867–7872. https://doi.org/10.1073/pnas.1402785111

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. https://doi.org/10.1523/JNEUROSCI.3411-11.2012

Cléry, H., Bonnet-Brilhault, F., Lenoir, P., Barthelemy, C., Bruneau, N., & Gomot, M. (2013). Atypical visual change processing in children with autism: An electrophysiological Study. *Psychophysiology*, *50*(3), 240–252. https://doi.org/10.1111/psyp.12006

Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(7), 4002–4006. https://doi.org/10.1073/pnas.95.7.4002

Crane, L., Goddard, L., & Pring, L. (2009). Sensory processing in adults with autism spectrum disorders. *Autism*, *13*(3), 215–228. https://doi.org/10.1177/1362361309103794

Cross, D. V. (1973). Sequential dependencies and regression in psychophysical judgments. *Perception & Psychophysics*, *14*(3), 547–552. https://doi.org/10.3758/BF03211196

De Jonge, M. V., Kemner, C., & Van Engeland, H. (2006). Superior disembedding performance of high-functioning individuals with autism spectrum disorders and their parents: The need for subtle measures. *Journal of Autism and Developmental Disorders*, *36*(5), 677–683. https://doi.org/10.1007/s10803-006-0113-2

Del Viva, M. M., Igliozzi, R., Tancredi, R., & Brizzolara, D. (2006). Spatial and motion integration in children with autism. *Vision Research*, *46*(8–9), 1242–1252. https://doi.org/10.1016/j.visres.2005.10.018

Droit-Volet, S. (2003). Alerting attention and time perception in children. *Journal of Experimental Child Psychology*, *85*(4), 372–384. https://doi.org/10.1016/S0022-0965(03)00103-6

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. https://doi.org/10.1016/S0022-0965(02)00180-7

Droit-Volet, S., Clément, A., & Wearden, J. (2001). Temporal generalization in 3- to 8-year-old children. *Journal of Experimental Child Psychology*, *80*(3), 271–288. https://doi.org/10.1006/jecp.2001.2629

Droit-Volet, S., & Gautier, T. (2002). *Attentional distraction and time perception in children*.

Droit-Volet, S., Meck, W. H., & Penney, T. B. (2007). Sensory modality and time perception in children and adults. *Behavioural Processes*, *74*(2), 244–250. https://doi.org/10.1016/j.beproc.2006.09.012

Droit-Volet, S., & Wearden, J. H. (2001). Temporal bisection in children. *Journal of Experimental Child Psychology*, *80*(2), 142–159. https://doi.org/10.1006/jecp.2001.2631

Dyjas, O., Bausenhart, K. M., & Ulrich, R. (2012). Trial-by-trial updating of an internal reference in discrimination tasks: Evidence from effects of stimulus order and trial sequence. *Attention, Perception, and Psychophysics*, *74*(8), 1819–1841. https://doi.org/10.3758/s13414-012-0362-4

Ego, C., Bonhomme, L., Orban de Xivry, J. J., Da Fonseca, D., Lefèvre, P., Masson, G. S., & Deruelle, C. (2016). Behavioral characterization of prediction and internal models in adolescents with autistic spectrum disorders. *Neuropsychologia*, *91*, 335–345. https://doi.org/10.1016/j.neuropsychologia.2016.08.021

Ekman, G., Frankenhaeuser, M., Berglund, B., & Waszak, M. (1969). Apparent duration as a function of intensity of vibrotactile stimulation. *Perceptual and Motor Skills*, *28*(1), 151–156. https://doi.org/10.2466/pms.1969.28.1.151

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a. *Nature*, *415*(January), 429–433.

Espinosa-Fernández, L., de la Torre Vacas, L., García-Viedma, M. del R., García-Gutiérrez, A., & Torres Colmenero, C. J. (2004). Temporal performance in 4-8 year old children. The effect of chronometric information in task execution. *Acta Psychologica*, *117*(3), 295–312. https://doi.org/10.1016/j.actpsy.2004.08.001

Fabbri, M., Cancellieri, J., & Natale, V. (2012). The A Theory Of Magnitude (ATOM) model in temporal perception and reproduction tasks. *Acta Psychologica*, *139*(1), 111–123. https://doi.org/10.1016/j.actpsy.2011.09.006

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743. https://doi.org/10.1038/nn.3689

Frith, U., & Happé, F. (1994). Autism: beyond "theory of mind." *Cognition*, *50*(1–3), 115–132. https://doi.org/10.1016/0010-0277(94)90024-8

Fulcher, C. (2017). The role of sensory history and stimulus context in human time perception. Adaptive and integrative distortions of perceived duration. *Doctoroal Thesis, University of Bradford*, (Dcm). https://doi.org/10.1007/s40093-015-0101-7

Furmanski, C. S., & Engel, S. A. (2000). An oblique effect in human primary visual cortex. *Nature Neuroscience*, *3*(6), 535–536. https://doi.org/10.1038/75702

Gepner, B., Mestre, D., Masson, G., & de Schonen, S. (1995). Postural effects of motion vision in young autistic children. *NeuroReport*, *6*(8), 1211–1214. https://doi.org/10.1097/00001756-199505300-00034

Germanò, E., D, A. G. M., & Curatolo, P. (2012). Comorbidity of ADHD and Dyslexia Comorbidity of ADHD and Dyslexia. *Developmental Neuropsychology*, *5641*(June), 37–41. https://doi.org/10.1080/875656412010494748

Gibbon, Church, R, M, & M. (1984). *Scalar Timing in Memory* (Vol. 423).

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932. https://doi.org/10.1038/nn.2831

Glasauer, S., & Shi, Z. (2018). 150 years of research on Vierordt's law - Fechner's fault? *BioRxiv*, 450726. https://doi.org/10.1101/450726

Glasauer, S., & Shi, Z. (2019). *Central Tendency as Consequence of Experimental Protocol*. 268–271. https://doi.org/10.32470/ccn.2019.1148-0

Glasauer, S., & Shi, Z. (2021a). Individual believes about temporal continuity explain perceptual biases. *BioRxiv*, 1–21.

Glasauer, S., & Shi, Z. (2021b). The origin of Vierordt's law: The experimental protocol matters. *PsyCh Journal*, *10*(5), 732–741. https://doi.org/10.1002/pchj.464

Goldstone, S., Boardman, W. K., & Lhamon, William, T. (1958). *Kinesthetic cues in the development of time concepts*. 185.

Goldstone, S., & Lhamon, W. T. (1974). In human. *New York*, 63–82.

Goldstone, S., Lhamon, W. T., & Sechzer, J. (1978). Light intensity and judged duration. *Bulletin of the Psychonomic Society*, *12*(1), 83–84. https://doi.org/10.3758/BF03329633

Goldstone, S., & Smythe, Elizabeth, J. (1957). *The Time Sense: A Normative Genetic Study of the Development of Time Perception '*. 49–59.

Gomot, M., Giard, M. H., Adrien, J. L., Barthelemy, C., & Bruneau, N. (2002). Hypersensitivity to acoustic change in children with autism: Electrophysiological evidence of left frontal cortex dysfunctioning. *Psychophysiology*, *39*(5), 577–584. https://doi.org/10.1017/S0048577202394058

Green, D., Charman, T., Pickles, A., Chandler, S., Loucas, T., Simonoff, E., & Baird, G. (2009). Impairment in movement skills of children with autistic spectrum disorders. *Developmental Medicine and Child Neurology*, *51*(4), 311–316. https://doi.org/10.1111/j.1469-8749.2008.03242.x

Grondin, S. (2010). *Timing and time perception : A review of recent behavioral and neuroscience findings*. *72*(3), 561–582. https://doi.org/10.3758/APP

Grondin, Simon, Meilleur-Wells, G., Ouellette, C., & Macar, F. (1998). Sensory effects on judgments of short time-intervals. *Psychological Research*, *61*(4), 261–268. https://doi.org/10.1007/s004260050030

Hallez, Q., Damsma, A., Rhodes, D., van Rijn, H., & Droit-Volet, S. (2019). The dynamic effect of context on interval timing in children and adults. *Acta Psychologica*, *192*, 87–93. https://doi.org/10.1016/j.actpsy.2018.10.004

Happe, F. (1999). Autism: cognitive deficit or cognitive style? *Trends in Cognitive Sciences*, *3*(6), 216–222.

Happé, F., & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25. https://doi.org/10.1007/s10803-005-0039-0

Happé, F. G. E. (1997). Central coherence and theory of mind in autism: Reading homographs in context. *British Journal of Developmental Psychology*, *15*(1), 1–12. https://doi.org/10.1111/j.2044-835x.1997.tb00721.x

Happé, F. G., Mansour, H., Barrett, P., Brown, T., Abbott, P., & Charlton, R. A. (2016). Demographic and Cognitive Profile of Individuals Seeking a Diagnosis of Autism Spectrum Disorder in Adulthood. *Journal of Autism and Developmental Disorders*, *46*(11), 3469–3480. https://doi.org/10.1007/s10803-016-2886-2

Held, R., Birch, E., & Gwiazda, J. (1980). Stereoacuity of human infants. *Proceedings* *of the National Academy of Sciences of the United States of America*, *77*(9 II), 5572–5574. https://doi.org/10.1073/pnas.77.9.5572

Hellström, Å. (1979). Time errors and differential sensation weighting. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(3), 460–477. https://doi.org/10.1037/0096-1523.5.3.460

Hellström, Å. (1985). The Time-Order Error and Its Relatives. Mirrors of Cognitive Processes in Comparing. *Psychological Bulletin*, *97*(1), 35–61. https://doi.org/10.1037/0033-2909.97.1.35

Hellström, Å. (2000). Sensation weighting in comparison and discrimination of heaviness. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 6–17. https://doi.org/10.1037/0096-1523.26.1.6

Hellström, Å., Patching, G. R., & Rammsayer, T. H. (2020). Sensation weighting in duration discrimination: A univariate, multivariate, and varied-design study of presentation-order effects. *Attention, Perception, and Psychophysics*, 3196–3220. https://doi.org/10.3758/s13414-020-01999-z

Hellström, Å., & Rammsayer, T. H. (2015). Time-order errors and standard-position effects in duration discrimination: An experimental study and an analysis by the sensation-weighting model. *Attention, Perception, and Psychophysics*, *77*(7), 2409–2423. https://doi.org/10.3758/s13414-015-0946-x

Herrington, J. D., Baron-Cohen, S., Wheelwright, S. J., Singh, K. D., Bullmore, E. T., Brammer, M., & Williams, S. C. R. (2007). The role of MT+/V5 during biological motion perception in Asperger Syndrome: An fMRI study. *Research in Autism Spectrum Disorders*, *1*(1), 14–27. https://doi.org/10.1016/j.rasd.2006.07.002

Herry, C., Bach, D. R., Esposito, F., Di Salle, F., Perrig, W. J., Scheffler, K., … Seifritz, E. (2007). Processing of temporal unpredictability in human and animal amygdala. *Journal of Neuroscience*, *27*(22), 5958–5966. https://doi.org/10.1523/JNEUROSCI.5218-06.2007

Holland, M., K., & Lockhead G.,R. (1977). Sequential effects in absolute judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 92–104. https://doi.org/10.1037/0096-1523.3.1.92

Hollingworth, H., L. (1910). *The Central Tendency of Judgment*. *7*(17), 461–469.

Isaksson, S., S., S., Tuominen, J., Arstila, V., Falter-Vagner, C. M., & Noreika, V. (2018). *Is there a generalised timing impairment in Autism Spectrum Disorder* *across time scales and paradigms*.

Isaksson, Sofia, Salomäki, S., Tuominen, J., Arstila, V., Falter-Wagner, C. M., & Noreika, V. (2018). Is there a generalized timing impairment in Autism Spectrum Disorders across time scales and paradigms? *Journal of Psychiatric Research*, *99*(January), 111–121. https://doi.org/10.1016/j.jpsychires.2018.01.017

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*(21), 3621–3629. https://doi.org/10.1016/S0042-6989(99)00120-0

Jaffe-Dax, S., Raviv, O., Jacoby, N., Loewenstein, Y., & Ahissar, M. (2015). A computational model of implicit memory captures dyslexics' perceptual deficits. *Journal of Neuroscience*, *35*(35), 12116–12126. https://doi.org/10.1523/JNEUROSCI.1302-15.2015

Jamieson, D. G., & Petrusic, W. M. (1975). The dependence of time-order error direction on stimulus range. *Canadian Journal of Psychology*, *29*(3), 175–182. https://doi.org/10.1037/h0082023

Jarrold, C., Gilchrist, I. D., & Bender, A. (2005). Embedded figures detection in autism and typical development: Preliminary evidence of a double dissociation in relationships with visual search. *Developmental Science*, *8*(4), 344–351. https://doi.org/10.1111/j.1467-7687.2005.00422.x

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026. https://doi.org/10.1038/nn.2590

Jones, C. R. G., Swettenham, J., Charman, T., Marsden, A. J. S., Tregay, J., Baird, G., … Happé, F. (2011). No evidence for a fundamental visual motion processing deficit in adolescents with autism spectrum disorders. *Autism Research*, *4*(5), 347–357. https://doi.org/10.1002/aur.209

Karaminis, T., Cicchini, G. M., Neil, L., Cappagli, G., Aagten-Murphy, D., Burr, D., & Pellicano, E. (2016). Central tendency effects in time interval reproduction in autism. *Scientific Reports*, *6*, 1–13. https://doi.org/10.1038/srep28570

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the Absence of Clocks: Encoding Time in Neural Network States. *Neuron*, *53*(3), 427–438. https://doi.org/10.1016/j.neuron.2007.01.006

Karvelis, P., Seitz, A. R., Lawrie, S. M., & Seriès, P. (2018). Autistic traits, but not schizotypy, predict increased weighting of sensory information in bayesian visual integration. *ELife*, *7*, 1–29. https://doi.org/10.7554/eLife.34115

Kerns, K. A., McInerney, R. J., & Wilde, N. J. (2001). Time reproduction, working memory, and behavioral inhibition in children with ADHD. *Child Neuropsychology*, *7*(1), 21–31. https://doi.org/10.1076/chin.7.1.21.3149

Kerrigan, I. S., & Adams, W. J. (2013). Learning different light prior distributions for different contexts. *Cognition*, *127*(1), 99–104. https://doi.org/10.1016/j.cognition.2012.12.011

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object Perception as Bayesian Inference. *Annual Review of Psychology*, *55*(1), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Kiryakova, R. K., Aston, S., Beierholm, U. R., & Nardini, M. (2020). Bayesian transfer in a complex spatial localization task. *Journal of Vision*, *20*(6), 1–19. https://doi.org/10.1167/JOV.20.6.17

Kononowicz, T. W., Van Rijn, H., & Meck, W. H. (2016). Timing and Time Perception : A Critical Review of Neural Timing Signatures Before , During , and After the To-Be-Timed Interval . Timing and Time Perception : A Critical Review of Neural Timing. In *Sensation, Perception and Attention*.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*(9). https://doi.org/10.1371/journal.pone.0000943

Kumar, S. L. (2013). Examining the characteristics of visuospatial information processing in individuals with high-functioning Autism. *Yale Journal of Biology and Medicine*, *86*(2), 147–156.

Kups, A. (2021). *Type A and Type B E ects , Time-Order Error and Weber ' s Law in Human Timing Simulations and Synthesis Contents*. 1–36.

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*, *20*(9), 1293–1299. https://doi.org/10.1038/nn.4615

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. https://doi.org/10.1080/09541440802453006

Licari, M. K., Alvares, G. A., Varcin, K., Evans, K. L., Cleary, D., Reid, S. L., … Whitehouse, A. J. O. (2020). Prevalence of Motor Difficulties in Autism Spectrum Disorder: Analysis of a Population-Based Cohort. *Autism Research*, *13*(2), 298–

306. https://doi.org/10.1002/aur.2230

Lockyer, L., & Rutter, M. (1970). A five- to fifteen-year follow-up study of infantile psychosis. IV. Patterns of cognitive ability. *The British Journal of Social and Clinical Psychology*, *9*(2), 152–163. https://doi.org/10.1111/j.2044-8260.1970.tb00654.x

Lu, A., Hodges, B., Zhang, J., & Zhang, J. X. (2009). Contextual effects on number-time interaction. *Cognition*, *113*(1), 117–122. https://doi.org/10.1016/j.cognition.2009.07.001

Luman, M., Oosterlaan, J., & Sergeant, J. A. (2008). Modulation of response timing in adhd, effects of reinforcement valence and magnitude. *Journal of Abnormal Child Psychology*, *36*(3), 445–456. https://doi.org/10.1007/s10802-007-9190-8

Manning, C., Kilner, J., Neil, L., Karaminis, T., & Pellicano, E. (2017). Children on the autism spectrum update their behaviour in response to a volatile environment. *Developmental Science*, *20*(5), 1–13. https://doi.org/10.1111/desc.12435

Meck, W. H. (1996). Neuropharmacology of timing and time perception. *Cognitive Brain Research, 3*(3–4), 227-242.

Michels, W. C. ., & Helson, H. (1954). A Quantitative Theory of Time-Order Effects. *The American Journal of Psychology*, *67*(2), 327–334.

Milne, E., Swettenham, J., Hansen, P., Campbell, R., Jeffries, H., & Plaisted, K. (2002). High motion coherence thresholds in children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *43*(2), 255–263. https://doi.org/10.1111/1469-7610.00018

Milne, E., White, S., Campbell, R., Swettenham, J., Hansen, P., & Ramus, F. (2006). Motion and form coherence detection in autistic spectrum disorder: Relationship to motor control and 2:4 digit ratio. *Journal of Autism and Developmental Disorders*, *36*(2), 225–237. https://doi.org/10.1007/s10803-005-0052-3

Mitchell, P., & Ropar, D. (2004). Visuo-spatial Abilities in Autism: A Review. *Infant and Child Development*, *13*(3), 185–198. https://doi.org/10.1002/icd

Moore, D. G., Hobson, R. P., & Lee, A. (1997). Components of person perception: An investigation with autistic, non-autistic retarded and typically developing children and adolescents. *British Journal of Developmental Psychology*, *15*(4), 401–423. https://doi.org/10.1111/j.2044-835X.1997.tb00738.x

Mottron, L., Burack, J. A., Iarocci, G., Belleville, S., & Enns, J. T. (2003). Locally oriented perception with intact global processing among adolescents with high-functioning autism: Evidence from multiple paradigms. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *44*(6), 904–913. https://doi.org/10.1111/1469-7610.00174

Mottron, L., Dawson, M., & Soulières, I. (2009). Enhanced perception in savant syndrome: Patterns, structure and creativity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1522), 1385–1391. https://doi.org/10.1098/rstb.2008.0333

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Mullins, C., Bellgrove, M. A., Gill, M., & Robertson, I. H. (2005). Variability in time reproduction: Difference in ADHD combined and inattentive subtypes. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*(2), 169–176. https://doi.org/10.1097/00004583-200502000-00009

Narain, D., Mamassian, P., van Beers, R. J., Smeets, J. B. J., & Brenner, E. (2013). How the Statistics of Sequential Presentation Influence the Learning of Structure. *PLoS ONE*, *8*(4), 1–7. https://doi.org/10.1371/journal.pone.0062276

O'Riordan, M., & Plaisted, K. (2001). Enhanced discrimination in autism. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *54*(4), 961–979. https://doi.org/10.1080/713756000

Oprisan, S. A., & Buhusi, C. V. (2011). Modeling pharmacological clock and memory patterns of interval timing in a striatal beat-frequency model with realistic, noisy neurons. *Frontiers in Integrative Neuroscience*, *5*(September), 1–11. https://doi.org/10.3389/fnint.2011.00052

Ozonoff, S., Strayer, D. L., McMahon, W. M., & Filloux, F. (1994). Executive Function Abilities in Autism and Tourette Syndrome: An Information Processing Approach. *Journal of Child Psychology and Psychiatry*, *35*(6), 1015–1032. https://doi.org/10.1111/j.1469-7610.1994.tb01807.x

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin*, *143*(5), 521–542. https://doi.org/10.1037/bul0000097

Patching, G. R., Englund, M. P., & Hellström, Å. (2012). Time- and space-order effects in timed discrimination of brightness and size of paired visual stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 915–940. https://doi.org/10.1037/a0027593

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pellicano, E., & Burr, D. (2012). When the world becomes "too real": A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Pellicano, E., Gibson, L., Maybery, M., Durkin, K., & Badcock, D. R. (2005). Abnormal global processing along the dorsal visual pathway in autism: A possible mechanism for weak visuospatial coherence? *Neuropsychologia*, *43*(7), 1044–1053. https://doi.org/10.1016/j.neuropsychologia.2004.10.003

Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential Effects of Auditory and Visual Signals on Clock Speed and Temporal Memory. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1770–1787. https://doi.org/10.1037/0096-1523.26.6.1770

Petzschner, F. H. (2012). *A Bayesian Approach to Characteristic Behaviour in Path integration*.

Petzschner, 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. https://doi.org/10.1523/JNEUROSCI.2028-11.2011

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*(5), 285–293. https://doi.org/10.1016/j.tics.2015.03.002

Pfeuty, M., Ragot, R., & Pouthas, V. (2008). Brain activity during interval timing depends on sensory structure. *Brain Research*, *1204*(Umr 5231), 112–117. https://doi.org/10.1016/j.brainres.2008.01.022

Plaisted, K. C. (2015). Reduced generalization in autism: An alternative to Weak Central Coherence. *The Development of Autism: Perspectives from Theory and Research*, (2001), 149–169.

Plaisted, K., O'Riordan, M., & Baron-Cohen, S. (1998). Enhanced discrimination of novel, highly similar stimuli by adults with autism during a perceptual learning task. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(5), 765–775. https://doi.org/10.1017/S0021963098002601

Plaisted, K., Swettenham, J., & Rees, L. (1999). Children with autism show local precedence in a divided attention task and global precedence in a selective attention task. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*(5), 733–742. https://doi.org/10.1017/S0021963099004102

Powell, G., Meredith, Z., McMillin, R., & Freeman, T. C. A. (2016). Bayesian Models of Individual Differences: Combining Autistic Traits and Sensory Thresholds to Predict Motion Perception. *Psychological Science*, *27*(12), 1562–1572. https://doi.org/10.1177/0956797616665351

Price, K. J., Shiffrar, M., & Kerns, K. A. (2012). Movement perception and movement production in Asperger's Syndrome. *Research in Autism Spectrum Disorders*, *6*(1), 391–398. https://doi.org/10.1016/j.rasd.2011.06.013

Rakitin, B. C., Penney, T. B., Gibbon, J., Malapani, C., Hinton, S. C., & Meck, W. H. (1998). Scalar expectancy theory and peak-interval timing in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*(1), 15–33. https://doi.org/10.1037/0097-7403.24.1.15

Rammsayer, T. H., & Lima, S. D. (1991). Duration discrimination of filled and empty auditory intervals: Cognitive and perceptual factors. *Perception & Psychophysics*, *50*(6), 565–574. https://doi.org/10.3758/BF03207541

Rao, R. P., Olshausen, B. A., & Lewicki, M. S. (Eds.). (2002). Probabilistic Models of the Brain: Perception and Neural Function. *MIT Press*. https://doi.org/10.1086/377908

Rattat, A. C., & Droit-Volet, S. (2001). Variability in 5- and 8-year-olds' memory for duration: An interfering task in temporal bisection. *Behavioural Processes*, *55*(2), 81–91. https://doi.org/10.1016/S0376-6357(01)00168-1

Raviv, O., Ahissar, M., & Loewenstein, Y. (2012). How Recent History Affects Perception: The Normative Approach and Its Heuristic Approximation. *PLoS Computational Biology*, *8*(10). https://doi.org/10.1371/journal.pcbi.1002731

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 of the United States of America, *114*(2), 412–417. https://doi.org/10.1073/pnas.1610706114

Robertson, A. E., & Simmons, D. R. (2013). Robertson , A . E ., and Simmons , D . R . (2013) The relationship between sensory sensitivity and autistic traits in the general population . Journal of Autism and Developmental Disorders , 43 (4). pp . 775-784 . ISSN 0162-3257 Copyright © 2013 The A. *Journal of Autism and Developmental Disorders*, *43*(September), 775–784.

Robic, S., Sonié, S., Fonlupt, P., Henaff, M. A., Touil, N., Coricelli, G., … Schmitz, C. (2015). Decision-Making in a Changing World: A Study in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *45*(6), 1603–1613. https://doi.org/10.1007/s10803-014-2311-7

Ropar, D., & Mitchell, P. (2001). Susceptibility to illusions and performance on visuospatial tasks in individuals with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(4), 539–549. https://doi.org/10.1017/S002196300100717X

Sapey-Triomphe, L. A. (2017). Perceptual inference and learning in autism : a behavioral and neurophysiological approach. *Doctoral Dissertation*, *Université*.

Sapey-Triomphe, Laurie Anne, Timmermans, L., & Wagemans, J. (2021). Priors Bias Perceptual Decisions in Autism, But Are Less Flexibly Adjusted to the Context. *Autism Research*, *14*(6), 1134–1146. https://doi.org/10.1002/aur.2452

Sapey-Triomphe, Laurie Anne, Weilnhammer, V. A., & Wagemans, J. (2021). Associative learning under uncertainty in adults with autism: Intact learning of the cue-outcome contingency, but slower updating of priors. *Autism*. https://doi.org/10.1177/13623613211045026

Schab, F. R., & Crowder, R. G. (1988). The role of succession in temporal cognition: Is the time-order error a recency effect of memory? *Perception & Psychophysics*, *44*(3), 233–242. https://doi.org/10.3758/BF03206292

Shah, A., & Frith, U. (1983). an Islet of Ability in Autistic Children: a Research Note. *Journal of Child Psychology and Psychiatry*, *24*(4), 613–620. https://doi.org/10.1111/j.1469-7610.1983.tb00137.x

Shah, A., & Frith, U. (1993). Why Do Autistic Individuals Show Superior Performance on the Block Design Task? *Journal of Child Psychology and Psychiatry*, *34*(8), 1351–1364. https://doi.org/10.1111/j.1469-7610.1993.tb02095.x

Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences*, *17*(11), 556–564. https://doi.org/10.1016/j.tics.2013.09.009

Shimojo, S., Birch, E. E., Gwiazda, J., & Held, R. (1984). Development of vernier acuity in infants. *Vision Research*, *24*(7), 721–728. https://doi.org/10.1016/0042-6989(84)90213-X

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research*, *49*(22), 2705–2739. https://doi.org/10.1016/j.visres.2009.08.005

Skoczenski, A. M., & Norcia, A. M. (1999). Development of VEP vernier acuity and grating acuity in human infants. *Investigative Ophthalmology and Visual Science*, *40*(10), 2411–2417.

Smith, A., Taylor, E., Warner Rogers, J., Newman, S., & Rubia, K. (2002). *Evidence for a pure time perception deficit in children with ADHD*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/1469-7610.00043/full

Spencer, J., O'Brien, J., Riggs, K., Braddick, O., Atkinson, J., & Wattam-Bell, J. (2000). Motion processing in autism: Evidence for a dorsal stream deficiency. *NeuroReport*, *11*(12), 2765–2767. https://doi.org/10.1097/00001756-200008210-00031

Stocker, A. A., & Simoncelli, E. P. (2005). Constraining the prior and likelihood in a Bayesian model of human visual speed perception. *Journal of Vision*, *5*(8), 928–928. https://doi.org/10.1167/5.8.928

Stocker, Alan A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–585. https://doi.org/10.1038/nn1669

Stocker, Alan A., & Simoncelli, E. P. (2008). A Bayesian model of conditioned perception. *Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*.

Symons, L. A., Cuddy, F., & Humphrey, K. (2000). Orientation tuning of shape from shading. *Perception and Psychophysics*, *62*(3), 557–568. https://doi.org/10.3758/BF03212107

Takahashi, C. D., Nemet, D., Rose-Gottron, C. M., Larson, J. K., Cooper, D. M., & Reinkensmeyer, D. J. (2003). Neuromotor noise limits motor performance, but not motor adaptation, in children. *Journal of Neurophysiology*, *90*(2), 703–711. https://doi.org/10.1152/jn.01173.2002

Tassinari, H., Hudson, T. E., & Landy, M. S. (2006). Combining priors and noisy visual cues in a rapid pointing task. *Journal of Neuroscience*, *26*(40), 10154–10163. https://doi.org/10.1523/JNEUROSCI.2779-06.2006

Theisinger, L. A. (2022). *Time Processing and Predictive Coding in Autism Spectrum Disorder*. 1–80. Retrieved from https://edoc.ub.uni-muenchen.de/30803/7/Theisinger_Laura_Alena.pdf

Thomas, Ewart, A., C., & Weaver, Wander, B. (1975). Cognitive processing and time perception. *Proceedings of the IEEE/CPMT International Electronics Manufacturing Technology (IEMT) Symposium*, *17*(4), 363–367. https://doi.org/10.1109/iemt.2004.1321634

Thomas, E. C., & Brown, I. (1974). Time perception and the filled-duration illusion. *Perception & Psychophysics*, *16*(3), 449–458. https://doi.org/10.3758/BF03198571

Treisman, M. (1963). Temporal discrimination and the indifference interval. Implications for a model of the “internal clock”. *Psychological Monographs*, *77*(13), 1–31. https://doi.org/10.1037/h0093864

Van de Cruys, S., De-Wit, L., Evers, K., Boets, B., & Wagemans, J. (2013). Weak priors versus overfitting of predictions in autism: Reply to Pellicano and Burr (TICS, 2012). *I-Perception*, *4*(2), 95–97. https://doi.org/10.1068/i0580ic

Van de Cruys, S., Evers, K., Hallen, R. Van Der, Eylen, L. Van, Boets, B., & Wagemans, J. (2014). *Preise Minds in uncertain world*. *121*(4), 649–675.

Van de Cruys, S., Lemmens, L., Sapey-Triomphe, L. A., Chetverikov, A., Noens, I., & Wagemans, J. (2021). Structural and contextual priors affect visual search in children with and without autism. *Autism Research*, *14*(7), 1484–1495. https://doi.org/10.1002/aur.2511

Van de Cruys, S., Van der Hallen, R., & Wagemans, J. (2017). Disentangling signal and noise in autism spectrum disorder. *Brain and Cognition*, *112*, 78–83. https://doi.org/10.1016/j.bandc.2016.08.004

Van de Cruys, S., Vanmarcke, S., Van de Put, I., & Wagemans, J. (2018). The Use of Prior Knowledge for Perceptual Inference Is Preserved in ASD. *Clinical Psychological Science*, *6*(3), 382–393.

https://doi.org/10.1177/2167702617740955

Van der Hallen, R., Manning, C., Evers, K., & Wagemans, J. (2019). Global Motion Perception in Autism Spectrum Disorder: A Meta-Analysis. *Journal of Autism and Developmental Disorders*, *49*(12), 4901–4918. https://doi.org/10.1007/s10803-019-04194-8

Venter, A., Lord, C., & Schopler, E. (1992). A Follow-Up Study of High-Functioning Autistic Children. *Journal of Child Psychology and Psychiatry*, *33*(3), 489–597. https://doi.org/10.1111/j.1469-7610.1992.tb00887.x

Verstynen, T., & Sabes, P. N. (2011). How each movement changes the next: An experimental and theoretical study of fast adaptive priors in reaching. *Journal of Neuroscience*, *31*(27), 10050–10059. https://doi.org/10.1523/JNEUROSCI.6525-10.2011

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin and Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Wang, L. A. L., Petrulla, V., Zampella, C. J., Waller, R., & Schultz, R. T. (2022). Gross Motor Impairment and Its Relation to Social Skills in Autism Spectrum Disorder: A Systematic Review and Two Meta-Analyses. *Psychological Bulletin*, *148*(3–4), 273–300. https://doi.org/10.1037/bul0000358

Warren, D. H., Welch, R. B., & Mccarthy, T. J. (1981). *The role of visual-auditory " compellingness " in the ventriloquism effect : Implications for transitivity among the spatial senses*. *30*(6), 557–564.

Wearden, J. H. (1991). Do humans possess an internal clock with scalar timing properties? *Learning and Motivation*, *22*(1–2), 59–83. https://doi.org/10.1016/0023-9690(91)90017-3

Wearden, J. H., & Lejeune, H. (2008). Scalar properties in human timing: Conformity and violations. *Quarterly Journal of Experimental Psychology*, *61*(4), 569–587. https://doi.org/10.1080/17470210701282576

Wearden, John H. (2003). Chapter 2: Applying the scalar timing model to human time psychology: Progress and challenges. *Time and Mind II: Information Processing Perspectives*, 21–39.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604. https://doi.org/10.1038/nn858

Welch, R. B., DutionHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics*, *39*(4), 294–300. https://doi.org/10.3758/BF03204939

Woodrow, H. (1933). *Individual Differences in the Reproduction of Temporal Intervals Author (s): Herbert Woodrow Source : The American Journal of Psychology , Apr ., 1933 , Vol . 45 , No . 2 (Apr ., 1933), Published by : University of Illinois Press Stable URL : https://. 45*(2), 271–281.

Yan, J. H., Thomas, J. R., Thomas, K. T., & Stelmach, G. E. (2000). Developmental features of rapid aiming arm movements across the lifespan. *Journal of Motor Behavior*, *32*(2), 121–140. https://doi.org/10.1080/00222890009601365

Zakay, D., & Block, R. (1995). An attentional gate model of prospective time estimation. *Time and the Dynamic Control of Behavior*, (November 1994). Retrieved from http://www.montana.edu/wwwpy/Block/papers/Zakay%26Block-1995.pdf

Zanker, J., Mohn, G., Weber, U., Zeitler-Driess, K., & Fahle, M. (1992). The development of vernier acuity in human infants. *Vision Research*, *32*(8), 1557–1564. https://doi.org/10.1016/0042-6989(92)90211-Z

Zélanti, P. S., & Droit-Volet, S. (2011). Cognitive abilities explaining age-related changes in time perception of short and long durations. *Journal of Experimental Child Psychology*, *109*(2), 143–157. https://doi.org/10.1016/j.jecp.2011.01.003

Appendices

Appendix A

Repeated measures ANOVA for unfilled durations: Compression

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
Object	2.312e-4	1	2.312e-4	0.007	0.934	1.123e-4
Residuals	0.804	24	0.033			
Randomisation	0.388	1	0.388	27.779	< .001	0.189
Residuals	0.335	24	0.014			
Object ∗ Randomisation	0.005	1	0.005	0.245	0.625	0.003
Residuals	0.526	24	0.022			

Note. Type III Sum of Squares

Descriptives

Object	Randomisation	Mean	SD	N
Permanent	Walk	0.376	0.181	25
	Full	0.515	0.223	25
Intermittent	Walk	0.394	0.246	25
	Full	0.503	0.205	25

Appendix B

Repeated measures ANOVA for unfilled durations: CV

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
Object type	0.009	1	0.009	4.804	0.038	0.109
Residuals	0.045	24	0.002			
Randomisation	0.002	1	0.002	2.841	0.105	0.024
Residuals	0.017	24	6.883e-4			
Object type ✻ Randomisation	0.002	1	0.002	5.120	0.033	0.021
Residuals	0.008	24	3.482e-4			

Note. Type III Sum of Squares

Descriptives

Object type	Randomisation	Mean	SD	N
Permanent	Walk	0.224	0.072	25
	Full	0.224	0.069	25
Intermittent	Walk	0.213	0.063	25
	Full	0.196	0.053	25

Marginal Means

Marginal Means - Object type

		95% CI for Mean Difference		
Object type	Marginal Mean	Lower	Upper	SE
Permanent	0.224	0.198	0.249	0.013
Intermittent	0.205	0.179	0.230	0.013

Marginal Means - Object type ✻ Randomisation

			95% CI for Mean Difference		
Object type	Randomisation	Marginal Mean	Lower	Upper	SE
Permanent	Walk	0.224	0.198	0.250	0.013
Intermittent		0.213	0.187	0.240	0.013

Marginal Means - Object type ✻ Randomisation

			95% CI for Mean Difference		
Object type	Randomisation	Marginal Mean	Lower	Upper	SE
Permanent	Full	0.224	0.197	0.250	0.013
Intermittent		0.196	0.170	0.222	0.013

Simple Main Effects

Simple Main Effects - Object type

Level of Randomisation	Sum of Squares	df	Mean Square	F	p
Walk	0.001	1	0.001	1.009	0.325
Full	0.009	1	0.009	11.199	0.003

Note. Type III Sum of Squares

Appendix C

Repeated measures ANOVA for filled durations: Compression.

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
Object	0.173	1	0.173	9.646	0.003	0.063
Residuals	0.899	50	0.018			
Order	0.028	1	0.028	1.491	0.228	0.010
Residuals	0.943	50	0.019			
Object ∗ Order	7.466e-4	1	7.466e-4	0.054	0.818	2.725e-4
Residuals	0.697	50	0.014			

Note. Type III Sum of Squares

Descriptives

Object	Order	Mean	SD	N
Single	Walk	0.306	0.162	51
	Full	0.326	0.162	51
Multiple	Walk	0.244	0.166	51
	Full	0.272	0.159	51

Marginal Means - Object

		95% CI for Mean Difference		
Object	Marginal Mean	Lower	Upper	SE
Single	0.316	0.278	0.354	0.019
Multiple	0.258	0.220	0.296	0.019

Appendix D

Repeated measures ANOVA for filled durations: CV.

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
Object type	0.004	1	0.004	1.815	0.184	0.018
Residuals	0.108	48	0.002			
Randomisation	0.005	1	0.005	5.323	0.025	0.023
Residuals	0.047	48	9.688e-4			
Object type ∗ Randomisation	6.198e-4	1	6.198e-4	0.468	0.497	0.003
Residuals	0.064	48	0.001			

Note. Type III Sum of Squares

Descriptives

Object type	Randomisation	Mean	SD	N
Permanent	Walk	0.208	0.053	49
	Full	0.214	0.057	49
Intermittent	Walk	0.213	0.052	49
	Full	0.227	0.054	49

Marginal Means

Marginal Means - Randomisation

		95% CI for Mean Difference		
Randomisation	Marginal Mean	Lower	Upper	SE
Walk	0.210	0.198	0.223	0.006
Full	0.221	0.207	0.234	0.007

Appendix E

Chapter 4 Repeated Measures ANOVA: cue SD varied trials

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
cue SD	4.974	4	1.243	74.799	< .001	0.631
Residuals	1.197	72	0.017			
Order	0.033	1	0.033	1.026	0.325	0.004
Residuals	0.585	18	0.032			
cue SD ✻ Order	0.045	4	0.011	0.760	0.555	0.006
Residuals	1.055	72	0.015			

Note. Type III Sum of Squares

Descriptives

Descriptives

cue SD	Order	Mean	SD	N
0.01	Walk	0.012	0.058	19
	Full	0.027	0.055	19
0.05	Walk	0.056	0.049	19
	Full	0.107	0.118	19
0.10	Walk	0.179	0.113	19
	Full	0.210	0.149	19
0.15	Walk	0.327	0.187	19
	Full	0.389	0.252	19
0.20	Walk	0.460	0.191	19
	Full	0.435	0.173	19

Post Hoc Comparisons - cue SD

		Mean Difference	SE	t	p_{holm}
0.01	0.05	-0.063	0.030	-2.114	0.038

Post Hoc Comparisons - cue SD

		Mean Difference	SE	t	p_{holm}
	0.10	-0.175	0.030	-5.929	< .001
	0.15	-0.338	0.030	-11.441	< .001
	0.20	-0.429	0.030	-14.488	< .001
0.05	0.10	-0.113	0.030	-3.816	< .001
	0.15	-0.276	0.030	-9.327	< .001
	0.20	-0.366	0.030	-12.374	< .001
0.10	0.15	-0.163	0.030	-5.511	< .001
	0.20	-0.253	0.030	-8.558	< .001
0.15	0.20	-0.090	0.030	-3.047	0.006

Note. P-value adjusted for comparing a family of 10

Note. Results are averaged over the levels of: Order

Post Hoc Comparisons - Order

		Mean Difference	SE	t	p_{holm}
Walk	Full	-0.026	0.026	-1.013	0.325

Note. Results are averaged over the levels of: cue SD

Appendix F

Chapter 4 Repeated Measures ANOVA: walk SD varied trials.

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
walk SD	1.744	4	0.436	20.492	< .001	0.397
Residuals	1.532	72	0.021			
Order	0.022	1	0.022	0.986	0.334	0.005
Residuals	0.399	18	0.022			
walk SD ✻ Order	0.026	4	0.006	0.696	0.597	0.006
Residuals	0.667	72	0.009			

Note. Type III Sum of Squares

Descriptives

Descriptives

walk SD	Order	Mean	SD	N
0.01	Full	0.317	0.256	19
	Walk	0.336	0.288	19
0.05	Full	0.107	0.118	19
	Walk	0.056	0.049	19
0.10	Full	0.110	0.136	19
	Walk	0.074	0.061	19
0.15	Full	0.096	0.120	19
	Walk	0.074	0.071	19
0.20	Full	0.099	0.157	19
	Walk	0.080	0.052	19

Appendix G
Chapter 5 justification for cue SD and walk SD selection.

In this shortened version of the experiment, there was justification for which values of the cue SD and walk SD were selected out of the possible 10 combinations. In adult responses, the largest difference a between the size of compression was seen in conditions cue SD = 0.15 and walk SD = 0.05. We wanted to be confidence that a difference can be seen in the mean compression magnitude (1-b) between walk and full conditions, and due to children's increased amount of noise in responses (Droit Volet 2017), selecting a larger difference increased this possibility. Although a cue SD of 0.15 is fairly high. It is important for children that cue SD is small enough for them to recognise the relationship between the two coins, otherwise they may treat the yellow and red coin as having completely unrelated positions, especially if their ability to learn the standard deviation of coin locations is not fully formed (Chambers et al., 2018). Therefore, a cue and walk SD of 0.05 was selected as it showed the second largest difference in mean compression magnitude between the two stimulus condition but had a small enough cue SD to emphasise the relationship between the coins to the children.

Descriptives

Cue SD	Order	Mean (1-b)	SD (1-b)	N
0.01	Walk	0.012	0.058	19
	Full	0.027	0.055	19
0.05	Walk	0.056	0.049	19
	Full	0.107	0.118	19
0.10	Walk	0.179	0.113	19
	Full	0.210	0.149	19
0.15	Walk	0.327	0.187	19
	Full	0.389	0.252	19
0.20	Walk	0.460	0.191	19
	Full	0.435	0.173	19

Appendix H
Compression magnitude ANOVA

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
Randomisation	0.054	1	0.054	3.985	0.051	0.012
Randomisation ∗ Age	4.276e-4	2	2.138e-4	0.016	0.984	9.400e-5
Residuals	0.666	49	0.014			

Note. Type III Sum of Squares

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
Age	1.584	2	0.792	17.279	< .001	0.348
Residuals	2.245	49	0.046			

Note. Type III Sum of Squares

Descriptives

Randomisation	Age	Mean	SD	N
Full	6 to 8	0.387	0.264	18
	9 to 11	0.180	0.131	15
	Adults	0.107	0.118	19
Walk	6 to 8	0.346	0.208	18
	9 to 11	0.135	0.181	15
	Adults	0.056	0.049	19

Test for Equality of Variances (Levene's)

	F	df1	df2	p
Full	10.817	2	49	< .001
Random	10.211	2	49	< .001

Post Hoc Comparisons - Age

		Mean Difference	SE	t	p_{bonf}	p_{holm}
6, to, 8	9, to, 11	0.209	0.053	3.952	< .001	< .001
	Adults	0.285	0.050	5.720	< .001	< .001
9, to, 11	Adults	0.076	0.052	1.447	0.463	0.154

Note. P-value adjusted for comparing a family of 3

Note. Results are averaged over the levels of: Randomisation

Appendix I
ANOVA on the SD of residuals

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η^2
Order	3.221e-4	1	3.221e-4	0.463	0.499	6.863e-4
Order ∗ Age	0.001	2	5.299e-4	0.762	0.472	0.002
Residuals	0.034	49	6.959e-4			

Note. Type III Sum of Squares

Between Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η²
Age	0.252	2	0.126	33.869	< .001	0.536
Residuals	0.182	49	0.004			

Note. Type III Sum of Squares

Test for Equality of Variances (Levene's)

	F	df1	df2	p
Walk	10.304	2	49	< .001
Full	6.118	2	49	0.004

Descriptives

Order	Age	Mean	SD	N
Full	6-8 years	0.165	0.063	18
	9-11 years	0.107	0.049	15
	Adult	0.055	0.029	19
Walk	6-8 years	0.169	0.066	18
	9-11 years	0.102	0.039	15
	Adult	0.045	0.017	19

Appendix J
t-test for TOE D1 means for reproduction only.

Paired Samples T-Test

Measure 1		Measure 2	t	df	p
D1	-	D2	-1.979	6	0.048

Note. For all tests, the alternative hypothesis specifies that D1 is less than D2.

Note. Student's t-test.

Descriptives

	N	Mean	SD	SE
D1	7	0.621	0.248	0.094
D2	7	0.701	0.204	0.077

Appendix K
t-test for TOE CV for reproduction only.

Paired Samples T-Test

Measure 1		Measure 2	t	df	p
D1	-	D2	4.114	6	0.003

Note. For all tests, the alternative hypothesis specifies that D1 is greater than D2.

Note. Student's t-test.

Descriptives

	N	Mean	SD	SE
D1	7	0.243	0.026	0.010
D2	7	0.221	0.024	0.009

Appendix L
t-test for TOE D1 means for combined.

Paired Samples T-Test

Measure 1		Measure 2	t	df	p
D1	-	D2	-2.119	6	0.039

Note. For all tests, the alternative hypothesis specifies that D1 is less than D2.

Note. Student's t-test.

Descriptives

	N	Mean	SD	SE
D1	7	0.644	0.142	0.054
D2	7	0.773	0.253	0.096

Appendix M
t-test for TOE CV for combined.

Paired Samples T-Test

Measure 1		Measure 2	t	df	p
D1	-	D2	1.144	6	0.148

Note. For all tests, the alternative hypothesis specifies that D1 is greater than D2.

Note. Student's t-test.

Descriptives

	N	Mean	SD	SE
D1	7	0.194	0.018	0.007
D2	7	0.179	0.038	0.014

Appendix N

Individual Akaike weights from Chapter 2: Filled

Permanent walk	Permanent full	Intermittent walk	Intermittent full	Mean
0.00	0.46	1.00	0.91	0.59
0.01	1.00	0.99	1.00	0.75
0.95	0.26	0.32	1.00	0.63
1.00	1.00	1.00	0.94	0.98
0.00	0.00	0.00	0.14	0.04
0.67	0.01	0.99	0.37	0.51
1.00	0.89	0.00	1.00	0.72
1.00	0.08	0.99	0.70	0.69
0.94	0.42	0.99	0.21	0.64

1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	0.00	0.75
0.78	0.47	1.00	0.97	0.81
1.00	0.94	0.98	1.00	0.98
1.00	1.00	0.90	1.00	0.98
0.35	0.99	0.13	0.80	0.57
1.00	1.00	0.89	0.81	0.93
0.50	0.00	1.00	0.99	0.62
1.00	1.00	1.00	1.00	1.00
1.00	1.00	1.00	0.97	0.99
1.00	0.65	1.00	0.62	0.82
1.00	1.00	1.00	0.47	0.87
1.00	1.00	0.86	1.00	0.96
1.00	1.00	0.94	1.00	0.99
0.00	1.00	0.95	1.00	0.74
1.00	0.78	0.06	1.00	0.71
1.00	0.76	0.71	0.22	0.67
0.99	1.00	0.40	0.80	0.80
1.00	1.00	1.00	0.56	0.89
1.00	1.00	1.00	0.07	0.77
1.00	1.00	1.00	1.00	1.00
0.99	1.00	1.00	0.00	0.75
0.00	0.00	1.00	1.00	0.50
1.00	0.98	0.93	0.16	0.77
0.61	1.00	1.00	1.00	0.90
1.00	0.53	1.00	0.99	0.88
1.00	0.00	1.00	1.00	0.75
1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.01	1.00	0.75
0.45	0.51	1.00	1.00	0.74
0.74	0.29	1.00	0.80	0.70
0.05	1.00	0.00	1.00	0.51
0.99	1.00	0.00	1.00	0.75
1.00	1.00	1.00	0.63	0.91
1.00	1.00	0.83	1.00	0.96
0.86	0.97	0.00	0.73	0.64
0.54	1.00	0.26	1.00	0.70
1.00	1.00	1.00	1.00	1.00
0.00	0.55	0.30	0.51	0.34

Individual Akaike weights from Chapter 2: Unfilled

Permanent walk	Permanent full	Intermittent walk	Intermittent full	Mean
0.00	1.00	0.79	1.00	0.70
1.00	0.90	1.00	1.00	0.98
0.83	0.98	1.00	0.38	0.80
0.64	0.34	1.00	0.15	0.53
1.00	1.00	0.03	1.00	0.76
1.00	1.00	0.45	1.00	0.86
1.00	1.00	0.97	0.83	0.95
1.00	1.00	1.00	1.00	1.00
0.00	0.39	0.77	0.99	0.54
1.00	0.56	0.91	0.98	0.86
0.00	0.06	1.00	1.00	0.51
1.00	1.00	0.00	1.00	0.75
1.00	1.00	1.00	0.99	1.00
0.72	0.45	0.91	1.00	0.77
1.00	1.00	1.00	1.00	1.00
0.04	1.00	0.78	1.00	0.70
0.01	1.00	1.00	1.00	0.75
1.00	1.00	1.00	1.00	1.00
1.00	1.00	0.48	0.85	0.83
1.00	1.00	1.00	1.00	1.00
0.98	1.00	1.00	0.35	0.83
0.18	0.61	0.94	0.61	0.58
1.00	0.14	0.00	1.00	0.54
1.00	1.00	1.00	1.00	1.00
0.00	1.00	0.00	1.00	0.50

Individual Akaike weights from Chapter 4

Individual Akaike weights for all conditions (adults)

Participant	No prior	Previous target	Running mean
1	0	1	2.34055E-48
2	0	1	4.47245E-73
3	1.8116E-263	3.23536E-15	1
4	0	1	3.9569E-102
5	7.7495E-202	0.017460403	0.982539597
6	0	1	5.46921E-44
7	4.0026E-278	1	1.5302E-57
8	9.6632E-173	0.999999961	3.89446E-08
9	6.43481E-75	1	1.86201E-24
10	7.6693E-231	3.01195E-46	1
11	2.2944E-132	9.48245E-25	1
12	1.0422E-242	1	8.88342E-66
13	1.8712E-295	1	1.04501E-62

15	1.8951E-220	1.60779E-36	1
16	0	1	4.27111E-91
17	1.1754E-252	7.26939E-20	1
18	2.9541E-160	0.003163682	0.996836318
19	6.10807E-70	4.37765E-10	1
20	7.2775E-245	1	2.30365E-30

Individual Akaike weights for random walk conditions (adults)

Participant	No prior	Previous target	Running mean
1	2.9089E-279	1	5.02258E-91
2	0	1	0
3	1.095E-111	4.90361E-07	0.99999951
4	5.1706E-266	1	1.10326E-91
5	1.77027E-51	0.947632173	0.052367827
6	0	1	0
7	3.5829E-145	1	2.67471E-54
8	4.6401E-102	1	5.16631E-16
9	2.23033E-50	1	1.79356E-23
10	8.7686E-155	1	4.18239E-30
11	4.45995E-95	4.62474E-14	1
12	4.7535E-200	1	4.34489E-38
13	8.3468E-186	1	6.66777E-52
15	1.3235E-180	1	3.24464E-64

16	0	1	0
17	2.6239E-204	1	1.79242E-39
18	8.6367E-159	1	6.31264E-15
19	0.932505645	0.00025843	0.067235925
20	1.0727E-120	1	2.7337E-42

Individual Akaike weights for fully random conditions (adults)

Participant	No prior	Previous target	Running mean
1	1.5505E-261	4.72349E-23	1
2	9.043E-237	2.65085E-48	1
3	2.1276E-148	4.111E-08	0.999999959
4	0	1	1.67019E-48
5	4.0927E-164	5.53418E-30	1
6	2.0951E-256	3.81712E-70	1
7	9.093E-178	2.5332E-29	1
8	1.14395E-74	0.013783387	0.986216613

9	8.49542E-25	0.999999882	1.18338E-07
10	1.5239E-109	2.58929E-43	1
11	1.46272E-64	8.51847E-15	1
12	4.4003E-120	1.59842E-07	0.99999984
13	1.7789E-126	0.999999965	3.5232E-08
15	1.1061E-112	2.7486E-30	1
16	9.29806E-95	0.99998113	1.88705E-05
17	1.1369E-101	3.3374E-20	1
18	1.579E-38	0.999794632	0.000205368
19	6.03754E-70	4.35963E-16	1
20	2.1237E-152	5.17892E-19	1

Individual Akaike weights Chapter 5

Akaike weights for all conditions (children)

Participant	No prior	Previous target	Running mean
3116	1.51376E-05	0.999872595	0.000112267
6053	2.62954E-16	0.999999761	2.39032E-07
6087	0.000497731	0.908705636	0.090796633
6874	9.16445E-06	0.907867846	0.092122989
6916	1.97677E-19	0.999999998	2.0736E-09
6946	2.58356E-18	1	1.94118E-13
6995	0.000904899	0.000185457	0.998909643
6997	9.90193E-08	0.057746201	0.9422537
7007	9.76672E-06	0.991220081	0.008770152
7018	3.42589E-05	0.999951099	1.46425E-05
7020	0.786986461	0.106506769	0.106506769
7040	0.436598431	0.244320613	0.319080956
7060	1.21863E-08	0.856769361	0.143230627
7064	3.22958E-08	0.99768261	0.002317357
7078	0.002016516	0.534203975	0.463779509
7093	1.83026E-05	0.553822894	0.446158803
7111	1.02427E-08	3.75303E-06	0.999996237
7129	0.78698776	0.10650604	0.1065062
7130	7.44381E-12	0.999999981	1.94363E-08
7132	0.412989363	0.255762438	0.331248198
7149	0.203101328	0.43107143	0.365827242
7153	0.000114356	0.193042018	0.806843626
7157	1.12571E-08	2.78352E-08	0.999999961
7168	0.786987132	0.106506381	0.106506487
7183	0.422907743	0.410066997	0.167025259
7186	0.526230166	0.307869793	0.165900041

7189	2.57195E-05	0.999563829	0.000410452
7205	0.000293847	0.999598851	0.000107302
7206	0.001045578	0.997944787	0.001009635
7210	0.507640092	0.07032049	0.422039418
7219	2.80455E-10	0.012186616	0.987813384
7227	9.00701E-07	0.130889418	0.869109681
7245	0.000147389	0.000256262	0.999596348

Akaike weights: Random walk only

Participant	No prior	Previous target	Running mean
3116	0.425581775	0.488338305	0.08607992
6053	7.51853E-07	0.999745075	0.000254173
6087	0.024382051	0.791219593	0.184398356
6874	4.15219E-05	0.992956051	0.007002427
6916	9.7689E-07	0.999989603	9.41985E-06
6946	1.49734E-11	1	4.21449E-11
6995	0.458688769	0.319680135	0.221631096
6997	0.000130338	0.760120244	0.239749418
7007	0.05152024	0.905230397	0.043249363
7018	0.704191055	0.20050724	0.095301705
7020	0.786986503	0.106506748	0.106506748
7040	0.786986365	0.106506826	0.10650681
7060	0.025627955	0.314766991	0.659605054
7064	0.027628473	0.528293002	0.444078526
7078	3.09118E-05	0.994756903	0.005212185
7093	0.001317921	0.031823737	0.966858342
7111	0.054781872	0.025278868	0.919939259
7129	0.753838954	0.144173143	0.101987903

7130	0.175280719	0.772719529	0.051999752
7132	0.781989417	0.110990361	0.107020222
7149	0.715957696	0.097305375	0.186736929
7153	9.60525E-06	0.996293928	0.003696466
7157	1.05354E-06	9.18617E-07	0.999998028
7168	0.786986939	0.10650653	0.10650653
7183	0.290709622	0.610779985	0.098510393
7186	0.786986101	0.10650695	0.10650695
7189	0.001955358	0.996808749	0.001235893
7205	0.537417121	0.389851425	0.072731455
7206	0.785441788	0.106297859	0.108260352
7210	0.721440476	0.098689837	0.179869687
7219	1.30348E-13	0.999999705	2.95146E-07
7227	2.93994E-07	0.992854934	0.007144772
7245	0.26555849	0.081809933	0.652631577

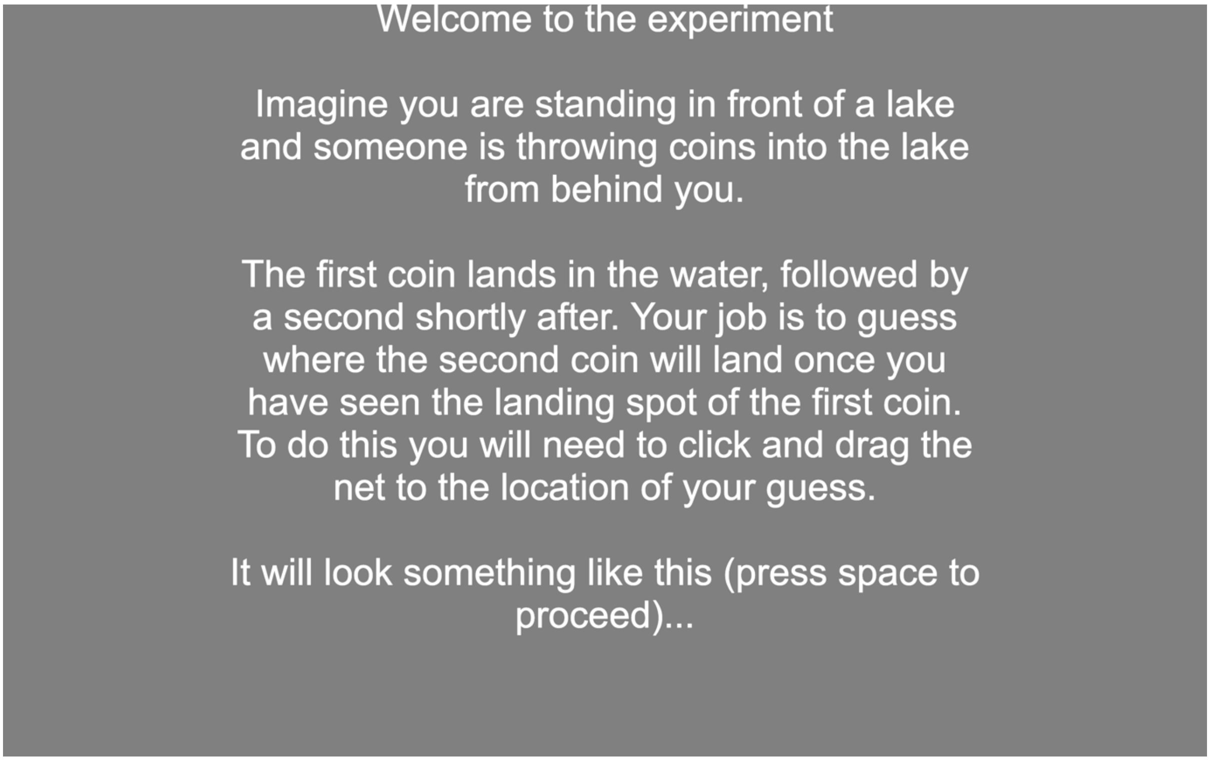
Akaike weights: Fully random only

Participant	No prior	Previous target	Running mean
3116	3.94046E-05	0.998728475	0.001232121
6053	2.15196E-09	0.997561216	0.002438782
6087	0.05921271	0.737431142	0.203356148
6874	0.045980583	0.665301515	0.288717902
6916	4.57006E-18	0.651359713	0.348640287
6946	1.9498E-08	0.998990006	0.001009974
6995	0.000366905	4.9655E-05	0.99958344
6997	0.000189558	0.021994273	0.977816169
7007	0.00141228	0.516299458	0.482288263
7018	1.21298E-06	0.994275875	0.005722912
7020	0.771873968	0.12366916	0.104456872
7040	0.011478884	0.312173723	0.676347393

7060	2.55753E-06	0.492066582	0.50793086
7064	8.11859E-06	0.992291245	0.007700636
7078	0.679460303	0.158431181	0.162108516
7093	0.007638796	0.896287948	0.096073256
7111	1.18145E-07	9.50731E-06	0.999990375
7129	0.786987245	0.106506375	0.10650638
7130	6.66923E-10	0.999982161	1.78382E-05
7132	0.004479435	0.023057328	0.972463237
7149	0.065946981	0.834805736	0.099247283
7153	0.115479815	0.116163387	0.768356798
7157	0.103429	0.089938607	0.806632393
7168	0.773748532	0.121538426	0.104713042
7183	0.734734893	0.139034791	0.126230316
7186	0.149772298	0.352930508	0.497297195
7189	0.027790252	0.852661516	0.119548232
7205	0.003175275	0.991785671	0.005039054
7206	2.71685E-05	0.999871671	0.000101161
7210	0.57012963	0.078392445	0.351477925
7219	0.000583393	0.00442693	0.994989677
7227	0.032860752	0.149806422	0.817332827
7245	0.000324515	0.000303274	0.999372211

Appendix O

Instructions for adult participants

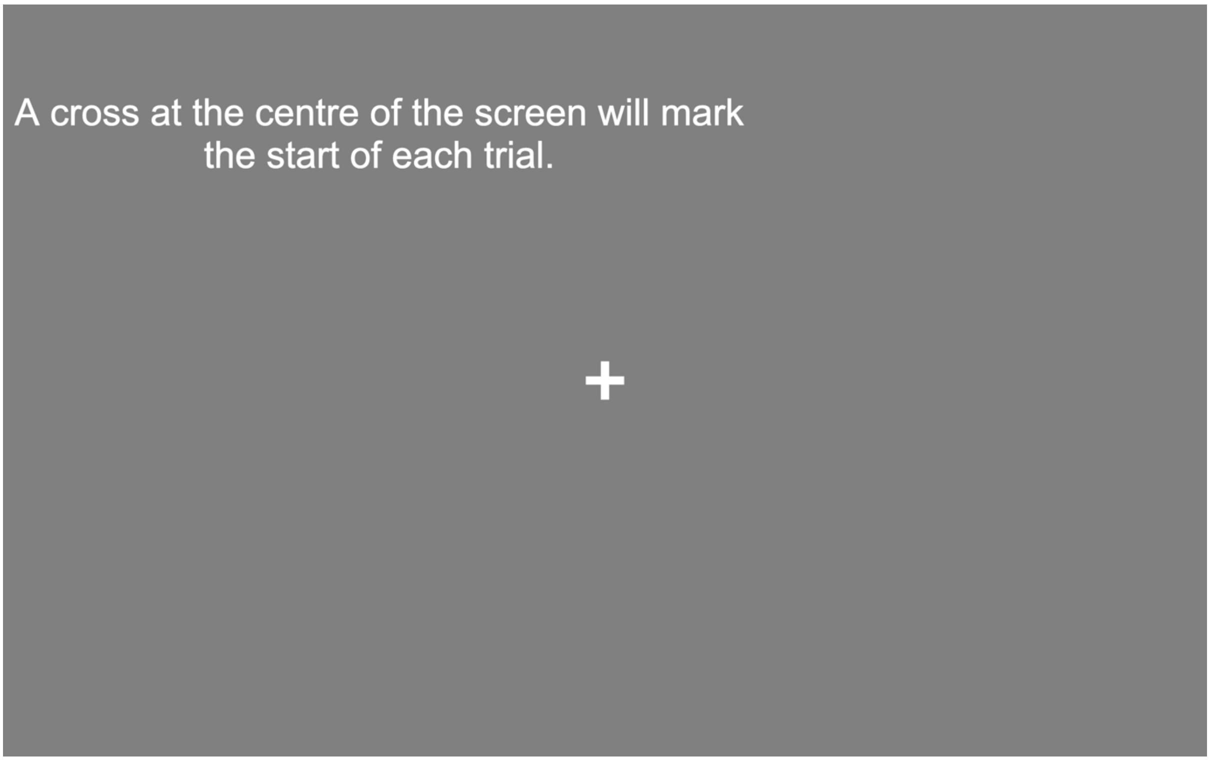


Welcome to the experiment

Imagine you are standing in front of a lake
and someone is throwing coins into the lake
from behind you.

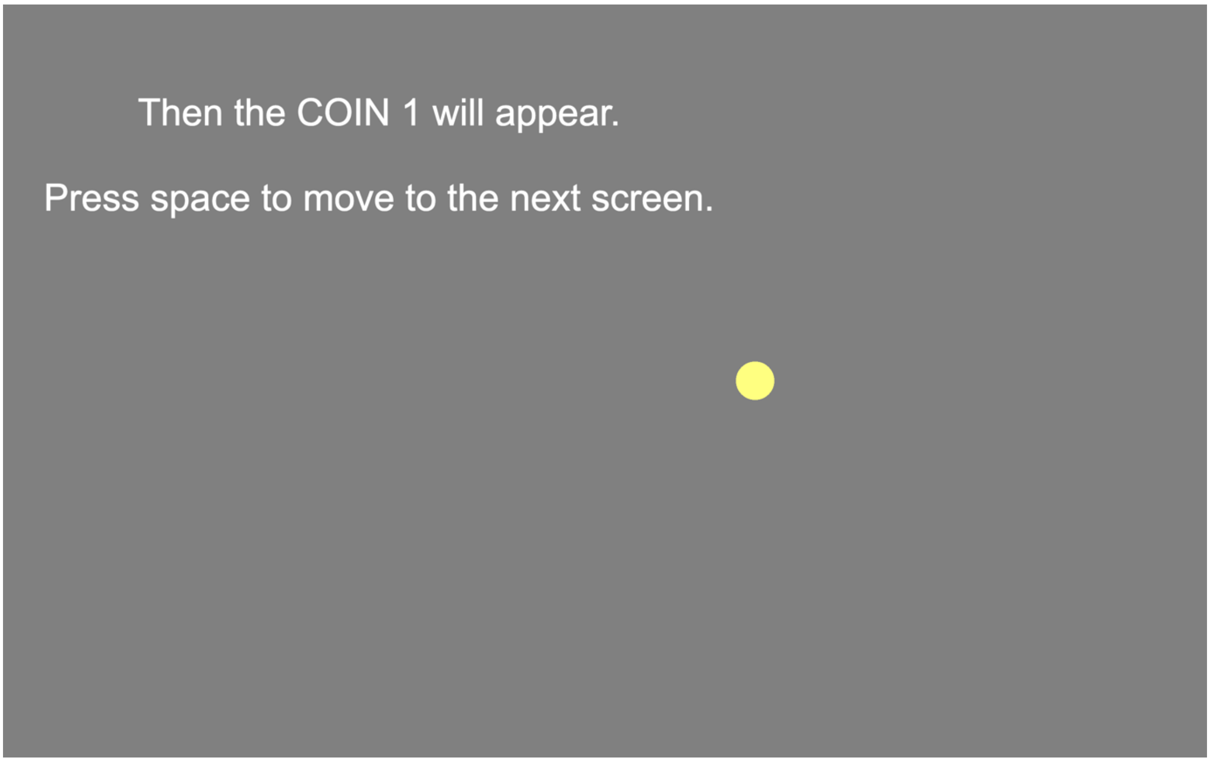
The first coin lands in the water, followed by
a second shortly after. Your job is to guess
where the second coin will land once you
have seen the landing spot of the first coin.
To do this you will need to click and drag the
net to the location of your guess.

It will look something like this (press space to
proceed)...



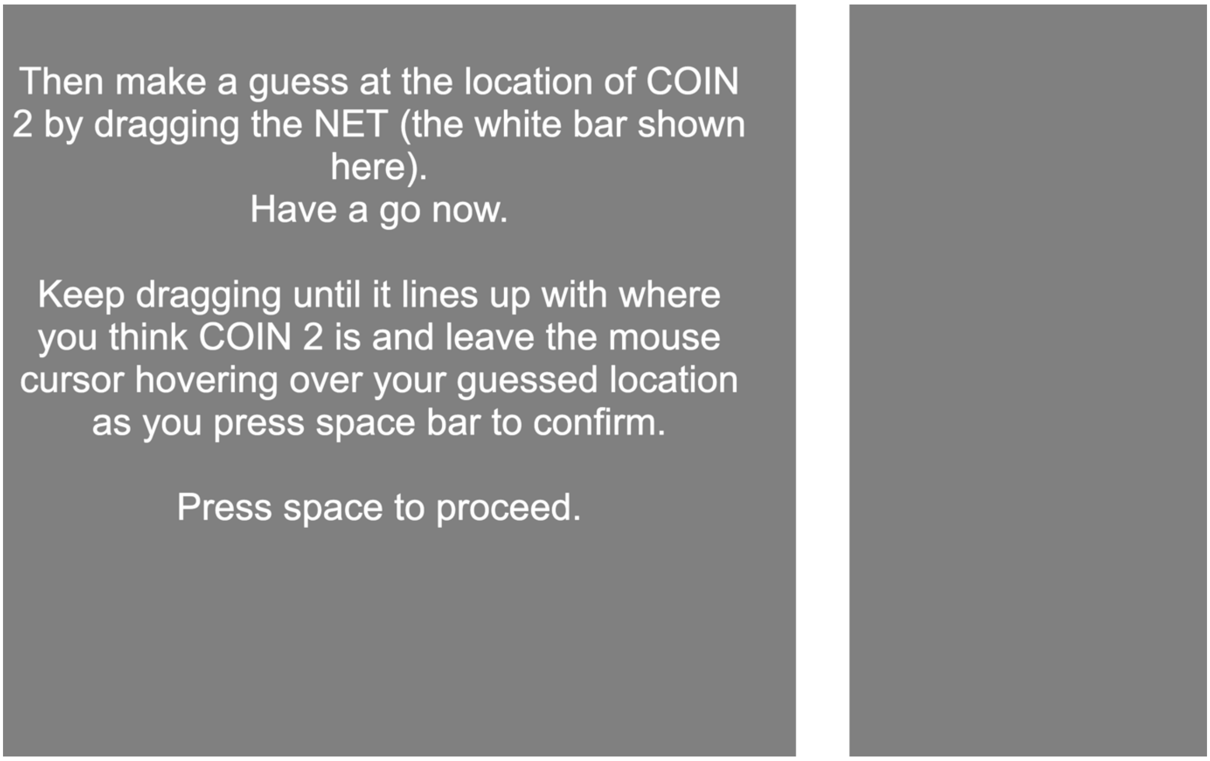
A cross at the centre of the screen will mark the start of each trial.

+



Then the COIN 1 will appear.

Press space to move to the next screen.



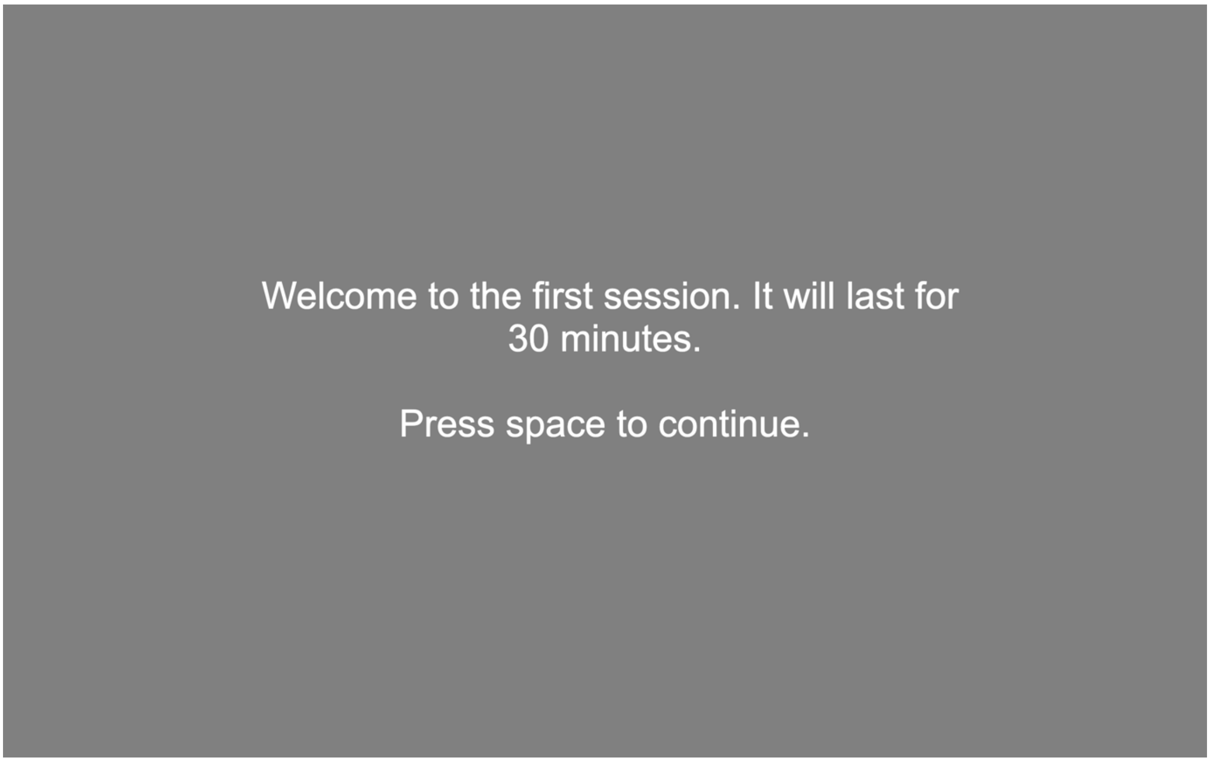
Then make a guess at the location of COIN 2 by dragging the NET (the white bar shown here).
Have a go now.

Keep dragging until it lines up with where you think COIN 2 is and leave the mouse cursor hovering over your guessed location as you press space bar to confirm.

Press space to proceed.

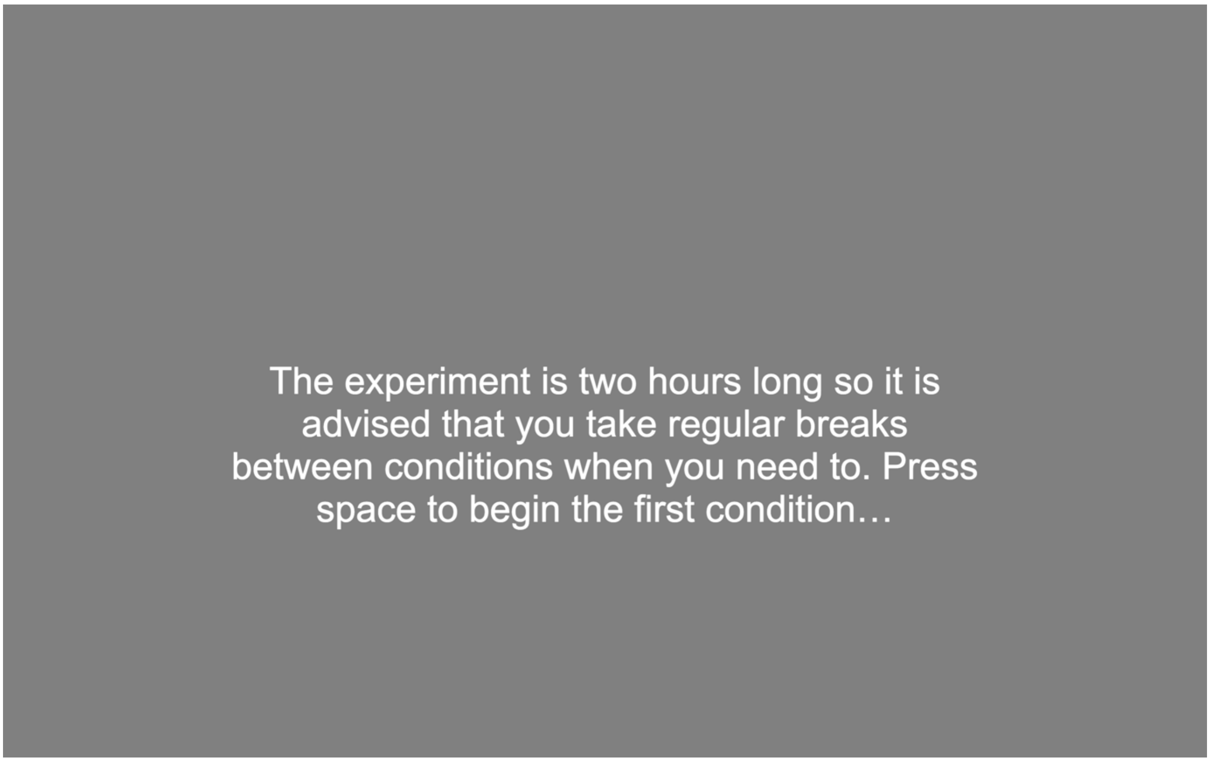
The location of the COIN 2 will then be revealed . Here it is!

Press space to begin the experiment.



Welcome to the first session. It will last for
30 minutes.

Press space to continue.



The experiment is two hours long so it is advised that you take regular breaks between conditions when you need to. Press space to begin the first condition…