Durham E-Theses

The development of perceptual priors

KIRYAKOVA, RENETA,KRASIMIROVA

How to cite:

KIRYAKOVA, RENETA,KRASIMIROVA (2021) *The development of perceptual priors*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/14211/

The development of perceptual priors

Reneta Kiryakova, BSc (Hons), MA.

Abstract

Bayesian inference has come to be regarded as the best, statistically optimal, way to deal with the sensory uncertainty inherent in our natural environment. One way to cope with such uncertainty is to incorporate our pre-existing knowledge about the world. However, we know very little about the circumstances in which human observers integrate sensory information with prior knowledge in a way that is close to optimal. We understand even less about how the developing brain adapts to the environmental statistics, learns to use them efficiently, and what factors may underlie the development of this critical perceptual skill. We addressed these questions though a series of psychophysical experiments, in which adults and 6- to 11-year-old children estimated the location of unseen targets based on a noisy sensory cue and a prior distribution that can be learned over the course of the experiment.

In Chapter 2, we showed that adult observers weighted sensory and prior information by their reliabilities but were far from optimal and struggled to generalise to untrained reliabilities in complex situations. The findings of Chapter 3 showed that 6- to 8-year-olds also weighted priors in proportion to their reliability, but they were slow to tune their behaviour to the statistics over time and remained furthest from optimal. Six- to -eight-year-olds' performance reached adult-like levels when the priors were explicitly shown. Conversely, when the decision rule was made more complex, 6- to 8-year-olds' abilities to distinguish between the priors broke down and adults' performance became more child-like. These findings prompted us to investigate whether individual differences, specifically in working memory, may predict performance in adults. The distance from optimal was not predicted by working memory capacity, beyond general cognitive abilities.

Together, these studies offer fresh insights into the capacity and limitations both adults and 6-11-year-old children have in learning and efficiently using novel environmental statistics.

The development of perceptual priors

Reneta Krasimirova Kiryakova, BSc (Hons), MA.

Thesis submitted for the degree of Doctor of Philosophy
Department of Psychology

Durham University, 2021

Table of Contents

List of Figures

Chapter 1
General Introduction

Chapter 2
Bayesian transfer in a complex spatial localization task

Chapter 3
Development of efficient adaptation to novel task statistics

Chapter 4
Individual differences in working memory capacity do not predict efficient adaptation to novel statistics

Appendices

List of Tables

Chapter 2
Bayesian transfer in a complex spatial localization task

Chapter 3
Development of efficient adaptation to novel task statistics

Chapter 4
Individual differences in working memory capacity do not predict efficient adaptation to novel statistics

Declaration

I, Reneta Kiryakova, confirm that, unless stated otherwise, all the work presented in this thesis is my own. This work was supported by a research grant from the North East Doctoral Training Centre (ES/J500082/1).

Publication Note

The entirety of Chapter 2 has been published:

Kiryakova, R. K., Aston, S., Beierholm, U. R., & Nardini, M. (2020). Bayesian transfer in a complex spatial localization task. *Journal of vision*, *20*(6), 17-17.

The entirety of Chapter 2 has been submitted for publication at *Developmental Science*:

Kiryakova, R. K., Aston, S., Beierholm, U. R., & Nardini, M. (submitted). Development of efficient adaptation to novel task statistics

Statement of Copyright

Acknowledgements

Here we are.

First and foremost, I would like to say a big thanks to my supervisors Professor Marko Nardini and Dr Ulrik Beierholm. Marko for his endless enthusiasm, curiosity, and forward thinking; there was always another question to ask, more that we could probe to get further insight. Ulrik for his calming presence, uplifting and encouraging attitude, and whose comments and questions, though deceptively simple, sparked much thought. You both passed along knowledge of relevant resources, shared experiences, views, and feedback and put a great deal of time into me and my experiments – and for that I am really grateful. I am enormously grateful to Dr Stacey Aston, whose support went beyond hands-on help with the code and analysis that went into the experiments in Chapter 2 and help with interpretation of everything that came after – I learned from you what you do not get explicitly taught, e.g., good coding practices, the importance of taking part in outreach initiatives. You are one of the most genuinely supportive and sincere people I know, and you have been an inspiration. Thank you to Dr James Negen who never failed to remind me that my work is valued and important. I am extremely grateful to Rachel Graham, Molly Reeve, Esther Yip and Ellie Nixon for help with collecting the data for Chapter 4.

The research in this thesis was only made possible by the funding I received from the ESRC and North East Doctoral Training Centre (ES/J500082/1), friends and officemates who ran my pilot experiments, and the children and families who took time out of their evenings and weekends to take part in my experiments.

Thank you to my family whose faith in me never wavered. Lastly, and perhaps most importantly, I would like to thank Tom – knowing that I have a constant source of support in you made the PhD journey so much easier.

Chapter 1

General Introduction

In order to make good, effective decisions in an uncertain, ever changing environment, we need to more optimally use all the information we have available to us. There are two crucial ways in which evidence can be optimally used: by integrating **(1)** information across different sensory cues, or **(2)** a sensory cue with prior knowledge about the statistical structure of previous events. Cue combination and how it develops has recently attracted more attention, but we know very little about how we learn to integrate prior and sensory information, what the limits of this learning are, how this learning changes over the course of childhood and what factors may underlie this ability. Is adults' ability to generalise learned priors to untrained cue reliabilities limited by task complexity? Are children slower to learn and use novel priors than adults? Could we improve children's ability to learn and use priors by relaxing working memory demands? Could we make their learning worse by adding more levels of sensory uncertainty? Can individual differences in working memory capacity explain some of the individual variation in prior integration abilities in adults? We address these questions in this thesis.

This introduction begins with an outline of the aims of the thesis. We then provide an overview of the Bayesian framework that is used to address the questions outlined above in this thesis. Next, we outline what is known about cue combination in order to serve as a benchmark for the second component of optimal inference: integration of prior knowledge and sensory information. We first describe how adults use multiple cues to improve perception in ambiguous situations and following this, outline the developmental course of cue combination. After, we summarise research on biases in visual perception in adults and children, before switching my focus to 'priors'. In discussing priors, we discuss two distinct types of priors: structural and contextual priors, first in adults, and then in children. At the end of this section, we highlight the gaps in our knowledge that will be addressed in this thesis. We then introduce the limiting evidence of individual differences in the detection of statistical regularities and working memory capacity, creating a knowledge gap between the two – we then outline how this gap is addressed in this thesis. Finally, we summarise my findings and describe how they contribute to the literature more broadly. We then discuss the practical implications of this work, its limitations and future directions before summarising the key conclusions from this thesis.

1.1 Thesis Aims

The aim of the work described in this thesis was to determine whether adults can perform Bayesian inference in more complex situations, what factors influence adults' abilities to efficiently combine prior and sensory information and how children's abilities to learn and use novel perceptual priors efficiently change as children get older, gain experience and undergo changes in cognitive functions. We approached this question by asking adults and children, aged between 6 and 11 years to learn novel prior distributions, and to integrate them with sensory input to more accurately estimate the location of unseen objects.

1.2 Perception Can be Understood as a Process of Bayesian Inference

We can trace the beginnings of Bayesian theory all the way back to Helmholtz's work on "unconscious inference"; that the brain deals with the inherently ambiguous information from our senses by incorporating prior knowledge of the environment (Geisler & Kersten, 2002; Helmholtz, 1925; Kersten et al., 2004). Several well-known visual illusions nicely illustrate this (Gregory, 1997; Weiss et al., 2002). For example, when we see the four 'Pac-Man' shapes in Figure 1.1a, we perceive an illusory square. We must have inferred, based on previous experiences, that by far the most likely reason for such a percept is that there are four black circles with a white square on top. Also, the larger context can help us make sense of smaller details. For example, the same character can be perceived as the letter 'B' or the number '13', depending on whether you are focusing on the letters presented horizontally ('A' and 'C') or the numbers presented vertically ('12' and '14'; Figure 1.1b). Bayesian Decision Theory (BDT) proposes a normative framework for optimally interpreting such ambiguous retinal images (Kersten et al., 2004; Knill & Pouget, 2004; Knill & Richards, 1996; Maloney, 2002; Mamassian et al., 2002).

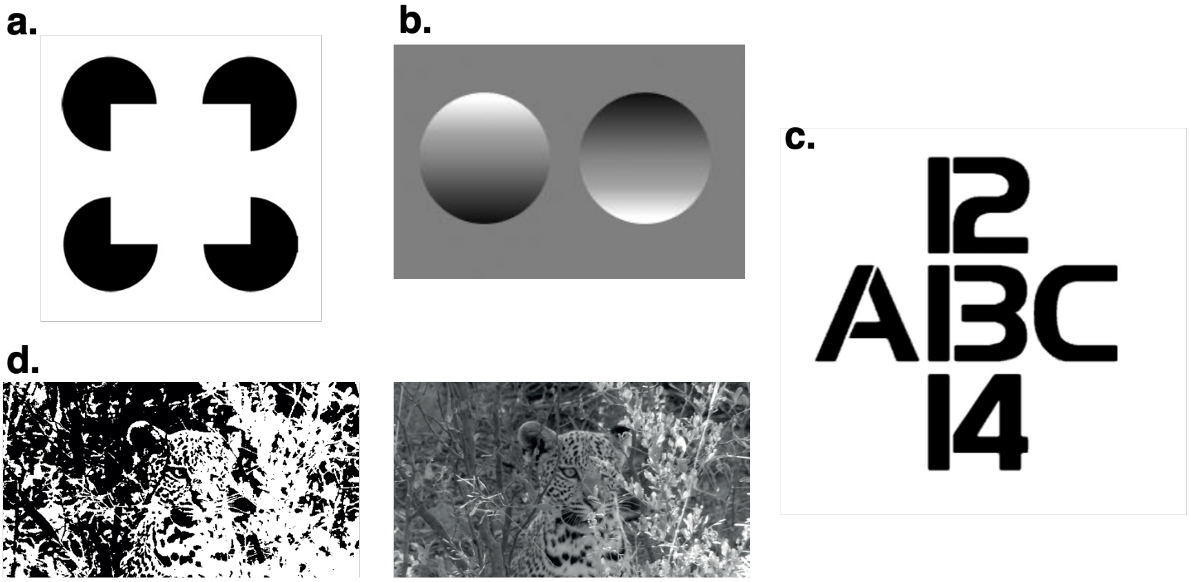

Figure 1.1. Perceptual illusions. **a.** A Kanizsa square: is that four ‘Pac-Man’ figures or four black circles with a white square overlaid on top? **b.** Surfaces that are brighter at the top are generally seen as convex and the others as concave, consistent with an assumption of light from above. **c.** Is the character in the centre the letter ‘B’ or the number ‘13’? **d.** Knowing that you are looking for a leopard makes the leopard features in the first image stand out more.

BDT has two key components: first, the hypothesis about the world. In its simplest form, this component can take the form of a prior distribution (Yuille & Kersten, 2006), but is more generally represented by a generative model (Clark, 2013). We will only be using the term ‘prior’ distributions in the following. The second component is the data received from the senses. These data, too, are known by different names, such as likelihood (Yuille & Kersten, 2006) or sensory information (Clark, 2013). Both terms are used interchangeably throughout this thesis.

When perception is described as a form of Bayesian inference, the prior and the likelihood are probability distributions which, together, influence what we perceive. Imagine playing tennis and doing your best to figure out where your opponent is likely to serve the ball (Figure 1.2). When playing an opponent for the first time, the ‘prior’ is a flat probability distribution: your opponent is equally likely to direct their serve either to the middle or to the side of the court. Suppose, over time, you observe more and more balls landing in the same vicinity (following a normal distribution with some random noise; red, Figure 1.2). Knowledge of your opponent's previous landing positions could then be used to predict the trajectory of the ball, even before it is hit. Combining this prior with the noisy information from your eyes (blue, Figure 1.2) via Bayes' rule gives a posterior probability (orange, Figure 1.2) that the ball is going to land in a position x:

$$P(x|o) = cP(o|x)P(x) \tag{1}$$

where $P(o|x)$ describes the uncertainty in the sensory information (the likelihood), or the probability of different observations given the true world state, x; $P(x)$ represents the prior belief over possible world states (the prior); and c is a normalisation constant.

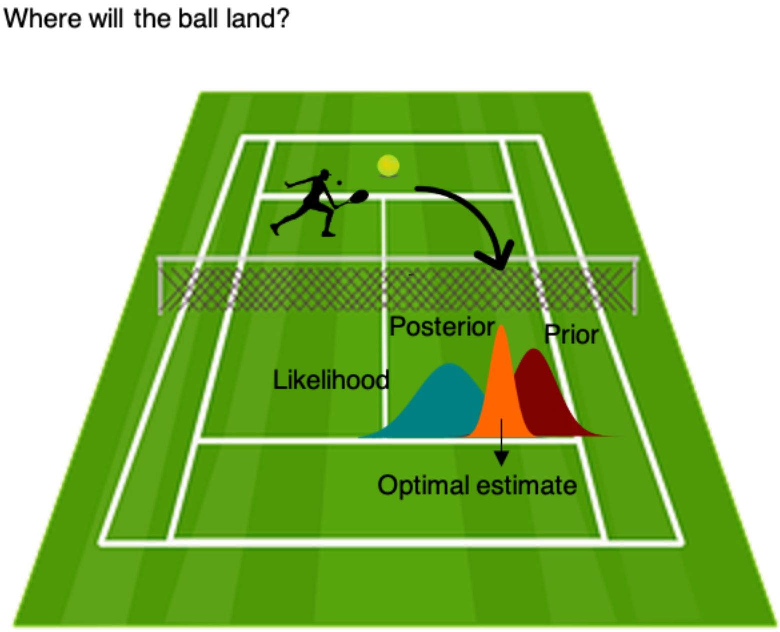

Figure 1.2. Example of integration of prior expectations and sensory evidence. **a.** When playing tennis, people can benefit from combining what they know from before (prior, red) with new sensory information (likelihood, blue). In doing so, they can estimate the posterior (red distribution) which helps to make the best, optimal decision.

A similar approach could be used to understand the perception of visual illusions – disambiguation of the central character (as a ‘B’ or ‘13’) in Figure 1.1c is an example. The sensory information provided by the character in the middle – its shape – matches the number ‘13’ and the letter ‘B’ equally well. Thus, without prior information, you would be equally likely to see this as either ‘B’ or ‘13’. Having priors can heavily alter what you perceive. When reading the letters horizontally, we might expect to see the letter ‘B’. This prior would then shift the resulting posterior distribution towards the letter ‘B’. If reading the numbers vertically, however, our prior would be for the number ‘13’.

As argued by Geisler and Kersten (2002), using Bayesian methods brings many advantages. Bayesian methods can be used to provide a normative way to measure whether people are close to optimal in perceptual tasks (Geisler & Kersten, 2002). One can then consider the ecological constraints (e.g., limited time, capacity limitations in attention and memory), under which the task must be performed, and thus gain deeper understanding of the factors that could stop people reaching optimal levels. Furthermore, ideal observer predictions are critical in providing benchmarks against which to compare human performance (Gardner, 2019; Heng et al., 2020; Qamar et al., 2013).

Secondly, the Bayesian approach tells us how to combine multiple information sources when the reliability of each source varies (Geisler & Kersten, 2002). The reliability of the different sources can change; we should therefore adapt the way we weight each information source based on how reliable they are. In the tennis analogy, as uncertainty increases, when playing in fog or at dusk, the Bayesian ideal observer should shift more weight to the prior probability distribution (in this case, the most likely ball landing position), relative to the visual input.

We wish to emphasize that because the Bayesian framework generates testable predictions, it is often useful even in the absence of evidence for an integration strategy (Stine et al., 2020). In cases where there is a mismatch between Bayesian predictions and actual performance, one can then consider potential alternative strategies that observers could be adopting. This is done by manipulating the different components of the model. Doing this might reveal differences between not using a piece of information (e.g., the prior) at all and mis-weighting it that would have otherwise gone undetected. As this example illustrates, there are various ways in which people can fail to use the (Bayes-)optimal strategy and understanding what those ways are is useful to infer and understand some of the (perhaps non-integration) strategies observers may deploy during everyday perception and action (Stine et al., 2020).

Despite its apparent flexibility, the Bayesian account has been criticised on several grounds: one of the most important is that the computations underlying the behaviours we observe are not clear (Jones & Love, 2011). With reference to the three levels of analysis (i.e., computational, algorithmic and implementation) proposed by Marr, the Bayesian approach speaks to the computational level as it focuses on normative principles (Marr, 1982). Process models, on the other hand, aim to uncover the mechanisms that underlie the decision process (Luce, 1995); these speak to Marr's algorithmic level (Marr, 1982). However, many standard Bayesian models also address 'how questions' at the algorithmic level. Many hypotheses proposed to account for suboptimal decisions focus on algorithmic level features such as capacity limitations (in working memory, for example), imprecisions, or the inability to employ complex decision rules (Rahnev & Denison, 2018). Thus, understanding suboptimal decision behaviour requires that we account for process-level considerations. In the present thesis, we begin with a normative account of how decisions generalise to other contexts (Chapter 2) but as we examine sources of suboptimality in human probabilistic inference, we also address process-level considerations, such as working memory and task complexity (Chapters 3 and 4).

1.3 Cue Combination

There are two components of optimal Bayesian inference: **(1)** cue combination, i.e., collating multiple sensory cues, and **(2)** integration of current sensory input with prior knowledge about the statistical structure of previous events. I begin by discussing the former.

1.3.1 Bayesian Predictions

Before asking whether human observers can do what the Bayesian theory says, we need to set out the predictions of optimal Bayesian integration – we do so here. In a typical cue combination task, there are two cues c_1 and c_2, both of which provide information about the same stimulus, s. For example, s could be the location of a stimulus, c_1 could be a visual cue for the location, and c_2 could be an auditory cue. If we assume that these cues are independent and that they are Gaussian with variances σ_1^2 and σ_2^2, respectively, then the location of the stimulus as estimated by the visual and auditory cues together can be computed as follows:

$$s_c = w_1 s_1 + w_2 s_2 \tag{1}$$

where w_1 is the weight assigned to the visual estimate, c_1, w_2 is the weight assigned to the auditory estimate, c_2. The combined estimate s_c is considered optimal because the variance of the estimate given both cues together is lower than that of the individual cues (see Equation 2), thus offering maximum precision – this is, in fact, the criteria to judge whether optimal cue integration is achieved. This is because it can rule out alternative explanations, such as switching between the sensory signals – we know this because if observers switched between cues, performance would be worse than the best of the two cues (Alais & Burr, 2019). The reduction in combined variance (and consequent gain in precision) is maximal when the variances of the sensory signals are reasonably well matched in reliability.

$$\sigma_c^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{2}$$

Going back to Equation 1, cue weights depend on their relative reliabilities (i.e., the inverse of variance), as shown in Equation 3 below. For clarity, only the weight for the visual estimate c_1 is shown.

$$w_1 = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \tag{3}$$

By varying the variance of one or both sensory signals, therefore, we could determine the degree to which each is being relied on. For example, if the visual cue c_1 is a less reliable source of information, observers should place less relative weight on it, compared to the auditory cue c_2. The down-weighting of cues that are less reliable is also what the ‘ideal observer’ model would predict.

Note that within this framework, combining sensory cues is mathematically equivalent to combining a prior and a sensory cue – this is done by straightforward replacement of one of the cues with the prior. The equations in this section, therefore, hold when combining a prior and sensory cue and analogous conclusions can be derived. For that reason, Equation 3 is used to compute the weight that the ideal observer should assign to the sensory cue in all experiments reported in this thesis.

1.3.2 Do Adults Do What the Bayesian Theory Says in Cue Combination Tasks?

As discussed in Section 1.3.1, Bayesian accounts make two predictions: (1) observers would be able to form a more reliable (i.e., less variable) percept of the environment when they have two sources of information to rely on, rather than only one, and (2) when the reliability of the different sources of information is varied, the resulting percept would be shifted towards the more reliable of the two sources (Clark & Yuille, 1990). Human behaviour generally matches these predictions in many cue combination tasks. For example, Ernst and Banks (2002) had four participants estimate the height of bars by looking at them or touching them. They first measured discrimination performance for the individual cues – visual and haptic - separately. Presenting both cues simultaneously, they found that subjects weighted by reliability, with less weight assigned to the visual signal when a large amount of noise was added to it. However, as was argued by Arnold et al. (2019), of the four levels of visual reliability, only one produced a bimodal performance that was clearly distinguishable from that achieved by the better of the cues (Ernst & Banks, 2002). Another study had participants localise visual and auditory stimuli – brief 'blobs' of light or 'clicks' of sound – in space, and found that subjects weighted the sensory signals according to their sensory reliabilities (Alais & Burr, 2004). This ability – to combine cues optimally - has now been reported for a wide range of tasks and combinations of cues, such as motion and texture cues to judge depth (Jacobs, 1999), and visual and proprioceptive cues to judge hand position (van Beers et al., 1999). Overall, these findings have been taken as evidence that the brain does take into account information about the relative reliabilities of cues when making perceptual decisions. Also, this ability is not only a 'cross-modal' one (Hillis et al., 2004; Jacobs & Fine, 1999; Louw et al., 2007, Landy & Kojima, 2001), as within the visual modality alone, adult humans make optimal use of stereo and texture information to estimate slant (Knill & Saunders, 2003).

1.3.3 Sub-optimal Cue Combination?

We reiterate that the Bayesian model predicts that (1) the discrimination threshold when both cues are available should be lower than that of either cue alone, and (2) that the discrimination threshold when both cues are available should not be higher than optimal predictions (Clark & Yuille, 1990). While some studies find that observers' behaviour matches these predictions (e.g., Ernst & Banks, 2002; Fetsch et al., 2009; see Section 1.3.2), others do not (Battaglia et al., 2003; Drugowitsch et al., 2014; Meijer et al., 2019). The latter set of studies generally show discrimination thresholds with both cues available (combined cue condition) to be significantly higher than optimal. For example, Meijer et al. (2019) found that although subjects' discrimination thresholds for the combined condition were lower than the better of the unimodal conditions, they were significantly higher than those predicted for optimal cue combination (Battaglia et al., 2003, 2011), suggesting that while observers can benefit from combining cues, they do not achieve the optimal gain. Several studies have shown that deviations from optimal behaviour are not specific to audio-visual localisation (Battaglia et al., 2003), but extend to many other tasks and combinations of cues (Bentvelzen et al., 2009; Burr et al., 2009; Butler et al., 2010; Fetsch et al., 2009; Maiworm & Röder, 2011; Prsa et al., 2012; Rosas et al., 2005).

As argued by Jones (2016), one reason why people may mis-weight cues, leading to suboptimal behaviour, may be that they are biased towards a given sense and that makes them weight it more than they should (i.e., than is optimal) (Fetsch et al., 2009; Maiworm & Röder, 2011). In our lifetime, we learn that vision could help more precisely localise sounds or estimate how far an object is than other senses – it therefore seems possible that in a task with auditory and visual stimuli presented at the same time, localisation is systematically biased in the direction of the visual stimulus because experience has taught observers that vision is typically the more reliable sense for spatial localisation (Battaglia et al., 2003; Meijer et al., 2019; Talsma et al., 2010). Another possibility is that observers wrongly estimated how reliable the cues were (Jones, 2016; Knill & Saunders, 2003). Even if there is some knowledge of one cue being more (or less) reliable than the other, these cues cannot be weighted in an appropriate (statistically optimal way) if the observer does not know *exactly* how reliable each cue is (Drugowitsch et al., 2014). This means that if the reliabilities are estimated wrongly, the observer could still compute the posterior, but it would be different from the true one (Ma, 2019). This is known as 'model mismatch': the reliabilities that the observer uses are not the same as the imposed cue reliabilities (Beck et al., 2012; Ma, 2019). If there is model mismatch, then observers are using a Bayesian strategy, but are not doing so optimally (Ma, 2012, 2019).

1.3.4 No Cue Combination

Findings of performance with combined cues being no better than that with either cue alone have also been reported (Chen & Tyler, 2015; Drugowitsch et al., 2014; Landy & Kojima, 2001; Oruç et al., 2003). For example, using a visual/vestibular heading discrimination task, Drugowitsch et al. (2014) not only found that the threshold for the combined, multisensory condition was not significantly lower than either the visual-only or vestibular-only thresholds (clearly violating prediction 1; see Section 1.3.3), but also that the combined threshold was significantly greater than that predicted by the optimal model (violating prediction 2; Section 1.3.3). This level of suboptimality goes further than that described in Section 1.3.3 (which revealed that sometimes cues are integrated but not weighted according to their reliabilities) because it shows that in some cases, cues are not integrated at all (Rahnev & Denison, 2018).

1.3.5 At What Age Does This Ability to Combine Cues Develop?

The vast majority of developmental research on cue combination has found abilities to combine cues optimally to emerge only after 8 years of age (Gori et al., 2008, 2012; Nardini et al., 2010). For example, Gori et al. (2008) asked children between the ages of 5 and 10 years to make size and orientation judgments based on vision only, haptic only and both visual and haptic information; 5-year-olds relied only on one sense (haptic for judging size, vision for judging orientation), but by 8-10 years, children were able to combine visual and haptic information in a statistically optimal way. A similar result was found in another study by the same authors (Gori et al., 2012): in younger children (< 12 years), vision dominated audition in bisection tasks; only adults combined information across the senses in an optimal manner. Nardini et al. (2010) also found that children older than 12 years integrated disparity and texture information optimally to reduce their uncertainty in judging surface slant but younger children did not. The narrative that emerges would seem to be of optimal cue combination in children, older than 8-10 years, and a dominant sensory modality at younger ages.

This conclusion was confirmed by Nardini et al. (2008), who examined the ability of adults, 4-5 year-olds and 7-8-year-olds to return an object to where it was originally placed by relying only on visual landmarks, only on self-motion cues or both together, and found no sign of integration in children younger than 8 years of age. As was pointed out, the experiment was not designed to test for optimality, but we still feel that the results support previous findings: adults combine multiple cues efficiently, but children younger than 8 years do not, instead relying on a single sense (Ernst, 2008). Similarly, Dekker et al. (2015) could not test whether subjects combined evidence optimally, but found a lack of adult-like (reliability-weighted) integration of stereo and texture information when judging surface slant in children younger than 10 years. It is, however, worth noting that not all studies investigating sensory integration during development have found delayed development of cue combination ability in younger children (< 12 years) relative to adults. For example, Jovanovic and Drewing (2014) found similar multisensory integration abilities in 6-year-old children and adults. Additionally, Adams (2016) showed that the ability of children to optimally combine emerged around 10 years of age, but some basic integration abilities, such as weighting sensory signals by their reliability, seem to already exist by the age of 4. Other studies have shown younger children to be more capable of combining cues than adults. For example, Petrini et al. (2016) tested 10-11-year-olds and adults in a path reproduction task and found that children near-optimally combined information across visual and self-motion cues, resulting in lower variable error; adults' behaviour, on the other hand, was indicative of cue competition.

The key conclusion that all the studies agree on is that optimal cue combination only occurs in children older than 8 years (though see Negen et al., 2019; Rohlf et al., 2020). But the studies disagree in two crucial ways. Firstly, there was no consensus of what happens before optimal cue combination 'comes online', with some studies finding sensory dominance (Gori et al., 2008; Nardini et al., 2010) and others finding evidence of reliability weighting in children as young as 4 years (Adams, 2016). Secondly, there was disagreement between the studies on the exact age at which optimal cue combination does occur, with the age at which subjects optimally integrate multiple cues changing according to task demands and cue combinations (e.g., Dekker et al., 2015; Gori et al., 2008, 2012; Nardini et al., 2013; Petrini et al., 2014). For example, Petrini et al. (2014) reported a lack of audio-haptic integration even by 12 years. By contrast, in a later study by the same authors (Petrini et al., 2016), children were shown to combine visual and self-motion cues when navigating as early as 10 years, and thus earlier than suggested by Petrini et al. (2014). This suggests that combining information optimally across different sensory cues, but also different tasks may show different time courses.

1.3.6 Why Does Optimal Cue Combination Occur So Late?

1.3.6.1 Need for Experience*.* A potential reason, and quite an obvious one, is that somewhere between 8 and 10 years of experience with pairs of cues is needed before these cues can be combined optimally. However, Negen et al. (2018) conclusively ruled out this possibility by showing that after two hours of training, adults not only learned a new auditory cue but also combined it optimally with a visual cue. This is evidenced by the fact that subjects had lower variable error with both cues together compared to the best of the two cues. Thus, cue combination abilities seem to emerge as we get older, regardless the amount of experience we have.

1.3.6.2 Need for calibration. An alternative explanation and one that has become increasingly influential in recent years is that children's senses are still calibrating – as would be expected for a growing body – meaning that senses cannot be integrated (Gori, 2015; Gori et al., 2008). When we have grown up, our body stops growing and remains roughly the same size for the rest of our lives. However, our body grows rapidly between birth and adulthood (Ernst, 2008). As young children grow up, their arms and legs stretch, for example, and this inevitably brings along a difference in proprioceptive signals about limb positions and movement. At the same time, inter-pupillary distance - that effective binocular vision depends on – increases. The sensory system, therefore, needs to be continuously updated to account for changes in body size during childhood (Ernst, 2008). Since mismatches between the senses hint at the need for recalibration, it may be more beneficial for the developing brain to keep the sensory cues separate and use conflict between them to decide whether to integrate or re-calibrate them (Ernst, 2008). Indeed, calibration has long been presumed to be a precursor of combining cues optimally (Gori et al., 2008 but see Rohlf et al., 2020).

However, a recent study has challenged this idea, showing that cross-modal recalibration emerges later than the ability to combine cues (Rohlf et al., 2020). Five-to- eleven-year-olds and adults localised sounds from one of six speakers, uniformly spaced along a semi-circle. The source of the sound was indicated by naming the animal which was closest to the sound. In multisensory trials, a visual distractor cue appeared to the side of the speaker the sound came from. The ventriloquist effect (indicating audio-visual integration), that is the shift of sound locations towards the visual distractor, was found even in 5-year-old children. By contrast, recalibration to the audio-visual spatial discrepancies was found to first emerge around the age of 8 years. Rohlf et al.'s (2020) results provide strong evidence against the long-believed hypothesis that cross-modal recalibration is a prerequisite for multisensory integration.

1.3.6.3 ‘Causal inference’ problem. By comparing different computational models, Rohlf and colleagues (2020) were also able to show that the ventriloquist effect in both children and adults was best explained by a causal inference model, i.e., that takes into consideration uncertainty about whether the cues are from the same object (so should be integrated using reliability weighting) or different objects (so should be kept separate). Interestingly, Rohlf et al. (2020) found that the model parameters varied with age, with younger children having lower prior expectations of common source. These findings are particularly important when considering that without knowing if the two cues are from the same stimulus, they can neither be integrated nor re-calibrated (Ernst, 2008; Kording et al., 2007). So, it could be that 5-year-olds may be unsure of the underlying causal structure and are still learning to assign different sensory cues to the same object (Dekker & Lisi, 2020), rather than that they lack the capability to combine cues optimally.

Why might younger children be uncertain about whether sensory cues are from the same or different objects? It may be because bias may be too high to make cue combination viable. This hypothesis was supported by Negen et al. (2019) who showed that artificially correcting young children’s biases by providing feedback on single cue trials allowed them to combine cues. Together, these results show that what develops in childhood may not be the basic mechanism allowing children to integrate cues, but other factors, such as whether the cues originated from the same source (Dekker & Lisi, 2020).

1.3.6.4 Need to Lean Cue Reliabilities. To combine cues optimally, we need to know how reliable each cue is. A fourth potential reason for the late maturation of optimal cue combination, therefore, is that young children need to learn the reliability of each sensory cue. How are the reliabilities of the sensory cues learned? One might imagine that they are learned through life-long experience of interacting with the environment. However, this is an unlikely explanation because the amount of variance in each sense and the number of environmental scenarios is infinite. Another way to learn cue reliabilities from experience is based on the correlation between pairs of cues (Jacobs, 2002). It is thought that within each perceptual domain, there is a ‘primary’ cue (e.g., vision for orientation, haptics for size and so on) that cannot be learned or modified. All other cues are ‘benchmarked’ against this cue, such that if their estimates match those given by the primary cue, they too are considered reliable. It is therefore possible that younger children may still be learning to recognise a given sense as a trustworthy ‘benchmark’.

1.3.6.5 Neural Maturation. Lastly, it could also be that the neural mechanisms that support the reliability-weighted integration of cues are immature (Rahnev & Denison, 2018). Recent evidence has implicated divisive normalisation, whereby the activity of one neuron is modulated by neighbouring neurons (Carandini & Heeger, 2013), in the more precise computing of reliability-weighted averages (Ohshiro et al., 2011). Thus, one potential explanation for the later onset of reliability-based cue weighting could be that this normalisation is reduced in younger ages (< 8 years). To the best of my knowledge, no study has ever tested this hypothesis in young children. Still, it is known that symptoms of autism are linked to an increased ratio of neural excitation to inhibition, which would result in reduced divisive normalisation (Rosenberg et al., 2015; Rubenstein & Merzenich, 2003 but see also Van de Cruys et al., 2018). At the same time, impaired ability to integrate multisensory information in autism has also been reported (Baum et al., 2015). It is thus possible that low-level multisensory weighting processes, such as those that depend on divisive normalisation, are immature during early development. Current models also suggest that cue integration might depend not only on low-level processes (e.g., balance of excitation vs. inhibition) but also on massive developmental changes in the cortical structure. For example, it could be that pathways from modality-specific brain regions to supramodal network of regions dedicated to integrating sensory signals are still developing, resulting in immaturities in reliability-based cue weighting (Iliescu & Dannemiller, 2008; Nardini et al., 2008, 2010).

1.4 Biases Towards Previous Stimuli

As we saw in Section 1.3, observers integrate evidence near-optimally across different sensory cues. Information about previous stimuli has also been found to be integrated, leading to more precise estimates (Fritsche et al., 2020). In fact, such biases or adaptation towards previous stimuli have been suggested to reflect Bayes-like processes of incorporating information from the past to aid perceptual decisions (van Bergen & Jehee, 2019; Kalm & Norris, 2017). We will briefly summarise the research evidence for two types of biases below: central tendency and serial dependence.

1.4.1 Central tendency and serial dependence

Central tendency is where perceptual estimates gravitate to the mean of recently seen stimuli (Hollingworth, 1910). For example, when estimating uncertain time intervals, subjects biased their timing estimates towards the interval duration on previous trials (Jazayeri & Shadlen, 2010, 2015). Cicchini et al. (2014) also showed that when mapping numbers onto space, observers adapted their responses in response to previous trials. Central tendency has been reported for a wide range of stimulus types and tasks, such as estimates of line length (Ashourian & Loewenstein, 2011; Duffy et al., 2010; Huttenlocher et al., 2000), sweetness (Riskey et al., 1979), facial expressions (Corbin et al., 2017; Roberson et al., 2007), hue (Olkkonen et al., 2014; Olkkonen & Allred, 2014) and time intervals (Jamieson, 1977; Jazayeri & Shadlen, 2010; Ryan, 2011). It has been proposed that such biases towards stimuli encountered in the past may arise from an optimal strategy of incorporating prior information (bias) into perceptual decisions (Cicchini et al., 2018; Kalm & Norris, 2017; van Bergen & Jehee, 2019). Such a strategy is optimal when the environment is stable because it is expected that new sensory input will be similar to previous input – we can exploit information we have just learned to maximise the efficient processing of new sensory inputs. This is also what the Bayesian theory says we should do – when sensory evidence is uncertain and difficult to judge, we should integrate it with a prior prediction about the current stimulus based on what we know from previous trials. A few studies on colour perception show that people do this (Olkkonen et al., 2014); in a series of experiments, observers were presented with a ‘reference’ stimulus and, after a delay, a ‘test’ stimulus, and asked which appeared more ‘bluer’/ ‘yellower’. In their first experiment, Olkkonen et al. (2014) showed that estimates of hue were biased towards the mean of stimuli presented during the experiment. In a second experiment, Olkkonen et al. (2014) varied the delay between the reference and test stimuli, and showed that increased delay from reference to test increased the noise in the sensory signals. With increased noise, reliance on prior knowledge (about distributions of hues) also became higher. Across all experiments, Olkkonen et al. (2014) demonstrated that hue estimates were systematically biased towards the average hue, and that bias were larger when sensory uncertainty was higher. This bias has been previously implied to result from integrating prior information with noisy sensory information (Olkkonen & Allred, 2014 but see also Ashourian & Loewenstein, 2011; Jazayeri & Shadlen, 2010). Crucially, the finding of increased bias with increasing sensory uncertainty confirms the theoretical predictions of the Bayesian model.

Recent work has exposed another history bias – where choices in the current trial are influenced simply by the preceding trial; although note that studies report a dependence on stimuli further back in the stimulus sequence; for example, up to three trials back (Fischer & Whitney, 2014). This type of bias is treated separately in the literature from the central tendency bias (a bias towards the stimulus one trial back compared to a bias towards the mean of all stimuli within a series) and has been termed serial dependence (van Bergen & Jehee, 2019; Fischer & Whitney, 2014). As Aston and colleagues (2021) point out, serial dependence has been reported for a wide range of visual features, such as orientation (Cicchini et al., 2017; Fischer & Whitney, 2014; Fritsche et al., 2017), numerosity (Cicchini et al., 2014; Corbett et al., 2011; Fornaciai & Park, 2018), spatial location (Bliss et al., 2017; Manassi et al., 2018; Papadimitriou et al., 2015), identity of faces (Liberman et al., 2014), gender of faces (Taubert, Alais, et al., 2016) and attractiveness (Kondo et al., 2012; Taubert, Van Der Burg, et al., 2016; Xia et al., 2016).

The distinction between central tendency bias and serial dependence seems unclear. Mattar et al. (2016) sought to disentangle central tendency from serial dependence biases by exploring their respective timescales. The two effects had different time scales: serial dependence operated over a shorter timescale and central tendency over a longer timescale. The temporal pattern of these effects was well captured by a single mechanism which utilized a weighted mixture of previous stimuli, where the relative contributions of previous stimuli was determined by an exponential integration time constant: when the constant was high, stimuli were combined over a relatively long period of time (i.e., central tendency), and when the constant was low, the response on the current trial was very similar to the response on the immediately preceding trial (i.e., serial dependence). This suggests that the distinction between these two types of bias may be linked to them having different time constants. It has been proposed that such a mixture mechanism accounts for the fact that in our natural environment, statistical patterns do not form only between events which occur one after the other, but also those that happen over a much longer period (Fritsche et al., 2020). Such a mixture account, therefore, has the advantage of not only accounting for effect of the most recently experienced stimuli, but of capturing dependencies beyond what we had just seen (Fritsche et al., 2020; Kalm & Norris, 2018).

Why do observers use what they had learned on previous trials as a prior for their decisions? Urai et al. (2019) argued that this may seem unexpected at first glance, given that in most experiments, stimuli and outcomes are uncorrelated across trials (Abrahamyan et al., 2016). However, events and scenes in the natural environment are rarely independent from one another (Yu & Cohen, 2009). Thus, it was argued that people may wrongly assume that this is also true in laboratory experiments (Yu & Cohen, 2009). In support of this idea, people have been shown to take into account the expected rate of change of statistics (stable vs. unpredictable) and adjust their biases accordingly (Braun et al., 2018; Glaze et al., 2015; Kim et al., 2017), revealing that people have certain prior expectations about environmental stability. Together, these observations may explain the greater weighting given to past vs. new observations that is typically observed (Cicchini et al., 2014). However, it is important to point out that while assuming that consecutive stimuli are statistically related may be incorrect in serial dependence tasks, incorporating prior information from recent decisions is, in fact, a rational heuristic strategy in most everyday situations.

1.4.2.2 Central Tendency in the Developed Brain. We know that trial history alters the way in which adults interpret incoming sensory signals (Cicchini et al., 2012, 2014, 2017, 2018). Surprisingly, we know very little about the age at which such biases develop in childhood. Indeed, only two studies, to our knowledge, have examined the development of central tendency biases. In a line reproduction study, Sciutti et al. (2014) showed that just as adults, 7-, 8-, 10, 11- and 13-year-olds reproduced the same length as longer or shorter when previous line lengths were longer or shorter, respectively. Children showed substantially lower spatial resolution than adults but weighted prior knowledge about the mean with sensory signals in a constant manner across development. Similarly, Karaminis et al. (2016) asked 6-14-year-old children and adults to estimate the duration of time intervals and found that children as young as 6 biased their estimates towards the mean interval duration on previous trials. From these studies, one can only conclude that by the age of 6-7 years, children can implicitly learn the statistics of a stimulus set (across a range of domains) and incorporate these statistics into their decisions (Adams, 2016).

1.5 Priors

The studies reviewed in the previous section show that outcomes / stimuli on one trial can impact subsequent trials, leading to serial dependence and central tendency biases. This happens even though successive trials should be treated as independent (Lak et al., 2018; Mendonça et al., 2020). Next, we will examine studies where consecutive trials are derived from the same distribution (hereafter referred to as a prior). Priors can be structural, that is statistical regularities in the natural environment that humans learn over their lifetime (e.g., that light comes from above, Adams et al., 2004), or contextual – learnt over the course of an experimental session.

1.5.1 Structural Priors

1.5.1.1 Structural Priors in the Adult Brain. There are many regularities in our natural environment (Dong & Atick, 1995; Schwartz et al., 2007; Simoncelli & Olshausen, 2001); for instance, in natural visual scenes, objects tend to be static or move slowly (Sotiropoulos et al., 2011; Stocker & Simoncelli, 2006; Weiss et al., 2002; Welchman et al., 2008). Studies have shown that people tend to underestimate the speed of moving objects (Blakemore & Snowden, 1999; Pretto et al., 2012; Snowden et al., 1998; Stocker & Simoncelli, 2006; Stone & Thompson, 1992; Thompson, 1982; Weiss et al., 2002; Welchman et al., 2008) and moving sounds (Senna et al., 2015). This is exactly what we would expect of an ideal observer (Stocker & Simoncelli, 2006; Weiss et al., 2002) because in the natural environment, speed estimates are generally noisy and objects tend to move slowly, resulting in a prior for slower speeds (Pretto et al., 2012).

Bayesian methods can explain speed underestimation (Stocker & Simoncelli, 2006) and assumptions of light coming from above (Gerardin et al., 2007; Langer & Bulthoff, 2001; Mamassian & Landy, 2001; Sun & Perona, 1997). Moreover, as we show in this paragraph, both priors can change with training, suggesting that although people learn such biases over the course of their lifetime, short periods of training can make them malleable (Adams et al., 2004, 2010; Champion & Adams, 2007). After humans were exposed to 1.5 hrs of tactile training, they perceived the light direction as shifted by approximately 10° from the baseline prior. In a separate experiment, Adams et al. (2010) confirmed that receiving either haptic or visual feedback forced observers to re-interpret a shaded stimulus. Moreover, this recalibration of individuals' light priors lasted for several days, suggesting that the light-from-above prior can be modified, at least temporarily (Adams et al., 2010). Also, human adults have been shown capable of learning multiple priors related to different illuminations after only a few hours of visual-tactile training (Kerrigan & Adams, 2013). Furthermore, Morgenstern et al. (2011) found that the light from above prior could be overridden by other cues. Even the prior for slow speeds could be flexibly overridden: participants ceased to expect grating stimuli to move slowly after spending multiple sessions looking at gratings moving at a faster speed (Sotiropoulos et al., 2011). Overall, these findings suggest that people tend to use the statistical regularities in their natural environments, but that the extent to which they use these regularities can depend on the context.

1.5.1.2 The Development of Structural Priors. Studies examining whether children use regularities in the natural environment – in particular, the light-from-above prior - in the same way as adults have found mixed results. Croydon et al. (2017) found that adults and 7-14-year-old children use a light-from-above prior to a similar degree. Moreover, how the light-from-above prior was used did not change between 7 and 14 years of age, suggesting that this prior is already present in children younger than 7 years. Pickard-Jones et al. (2020) also found no differences in the light source direction that 5-, 7-, 9- and 11-year-olds assumed. Moreover, in these children, the assumed light direction was widely consistent with estimates obtained in adult participants (Andrews et al., 2013), suggesting the direction of the prior does not change with age. However, they excluded any child participants who did not show sensitivity to stimulus orientation (15 out of 19 5-year-olds and 8 out of 18 7-year-olds) from the analyses and were thus only able to demonstrate that the light-from-above prior was similar to adults in those children who were sensitive to the stimulus orientation. Other studies reported increased use of the light-from-above prior in older children and adults, compared to younger ages (Stone & Pascalis, 2010; Thomas et al., 2010). Stone and Pascalis (2010) tested children between the age of 4 and 10 years in a shape discrimination task and reported increased tendency to interpret stimuli as if light comes from above with increasing age. Moreover, Thomas et al. (2010) demonstrated that children between 4 and 12 years and adults tended to favour the convex over the concave shape for a shaded polo stimulus, and assume that it is lit from above. However, when there was a mismatch between shape and lighting direction, convexity was more commonly assumed amongst younger ages (< 6 years) and light-from-above by older groups. These findings suggest that both convexity and light-from-above priors develop in the first few years of life but the way these priors are weighted changes as the child develops: convexity biases dominate in younger ages; by contrast, illumination biases (e.g., light-from-above) are the dominant ones in late childhood and adulthood.

The age-dependent increases in the use of the light-from-above that were observed by Stone and Pascalis (2010) and Thomas et al. (2010) seem to suggest a developmental trajectory over a longer period of time, from age 7 to 12 years of age. An alternative explanation is that the ability to use the prior develops at different ages, which, at group level, looks like age-related changes (Pickard-Jones et al., 2020). This raises the obvious question of why some children develop this ability earlier than others. People apply what they have learned from past experiences; it is, therefore, possible that if past experiences vary between individuals, the specific age at which different individuals may begin to use these experiences may also differ. The role of experience may be difficult to decouple from brain development and maturation by only using natural stimuli because as people get older, along with more mature neural networks, they also acquire more experience. As we discuss in the next section, using novel stimuli, - that everybody, regardless of age, has the same amount of experience with - is necessary to provide a stronger test of the effects of prior experiences.

1.5.2 Contextual Priors

Having reviewed the evidence for structural priors – that is, natural long-term statistics – in both the adult and the developing brain, we shall now consider the evidence for shorter-lived novel statistics (also referred to as contextual priors), e.g., distributions of hidden targets. This is important to consider because in the natural world, there are both long-term statistics that do not change a great deal, such as light direction or speed of moving objects, and short-term ones that may be more context specific, e.g., distributions of return shots learned with one tennis player (Körding & Wolpert, 2004) might not generalise to another.

1.5.2.1 Contextual Priors in the Adult Brain. It has been suggested (Acerbi et al., 2012; Körding & Wolpert, 2004) that the acquisition of novel perceptual priors partially depends on the complexity of the prior distribution, with more resources (e.g., time, memory storage) needed to handle more complex distributions (Acerbi et al., 2012; Körding & Wolpert, 2004). As such, we will first discuss studies that use simple prior distributions before we move on to studies with complex prior distributions.

1.5.2.1.1 Simple Distributions. There is evidence that when the prior is a simple Gaussian, adult observers learn prior uncertainty over time and combine this knowledge with current sensory inputs as predicted by an ideal observer. Körding and Wolpert (2004) analysed the performance of human adults as they reached to a target. Midway through the movement, a visual cue indicating the current position of the finger was flashed very briefly (~100ms). This cue was a cloud of dots with varying amounts of noise added to it to vary its reliability. The cue was offset to the right side of the finger’s true position by a distance, chosen randomly from a normal distribution with a mean of 1cm and SD of 0.5 cm; participants were oblivious to the offset. The endpoint of the movement was shown at the end, but only of those trials in which no external noise was added. After 1,000 learning trials, on trials where the cue was not present, participants corrected for the mean offset of the cursor by pointing in the opposite direction to the offset by 1cm. When the cue was present, participants pointed between the feedback and the mean offset, weighted by their reliability in a statistically optimal way. It is worth mentioning that the authors did not independently and directly measure sensory uncertainty and so they could not determine how optimal participants were in this study. Even so, Körding and Wolpert (2004) were the first to show that human adults combine current and prior information in a way that is close to ideal.

Körding and Wolpert's (2004) work has since been extended by Tassinari, Hudson, and Landy (2006). They conducted a separate experiment to independently record observers’ motor and sensory noise. The study found that subjects’ behaviour was in qualitative agreement with a Bayesian observer model but their performance, on average, fell short of ideal due to sensory noise in estimating the centroid of the stimulus and motor variability during reaching.

1.5.2.1.2 Complex Distributions. Körding and Wolpert (2004) conducted a separate experiment using the same reaching task as before (see Section 1.5.2.1.1) but this time with a bimodal distribution. To enable participants to better learn the bimodal shape of the distribution, the authors gave feedback on all trials, did not add extra noise to the visual cue and doubled the total number of trials, such that there were now 2,000 training trials and 2,000 test trials. Using computational modelling, the authors inferred observers' estimated sensory uncertainty and the separation between the two Gaussian distributions. The distance between the two distributions was estimated to be 4.8cm (across all subjects), close to the true value of 4cm, indicating that subjects accurately inferred the bimodal shape of the distribution. However, learning occurred on the order of thousands of trials and large individual differences remained. This suggests that observers learnt the bimodal distribution, but at the cost of using more resources, becoming slower and more variable.

More recently, Chalk et al. (2010) had participants estimate the direction of motion of coherently moving dots, and found that they began to expect the dots to move in the direction(s) they had most frequently seen, as shown by a bias towards reporting those directions. When the dots were moving in the most frequent directions, participants spotted this faster and more accurately that other, more infrequent motion directions. On trials where there were no dots, expectations about the most likely motion direction led to participants still reporting seeing dots (i.e., hallucinations) and that these dots moved in the most frequently presented directions. These data can be well explained by a Bayesian model that combines a noisy estimate of the current direction of the dot cloud with a prior about the most likely motion direction (Chalk et al., 2010; Karvelis et al., 2018). This observation was confirmed when the Bayesian model outperformed all other models the authors tested. Together with Körding & Wolpert's (2004) work, this study supports the idea that humans can learn and compute with bimodal prior distributions.

Using the visual motion estimation task established in the previous study (Chalk et al., 2010), Gekas et al. (2013) examined whether participants could learn two prior distributions simultaneously, with different colours differentiating between the distributions. In one experiment, they used uniform and bimodal distributions, and in a second experiment, they used bimodal and trimodal distributions. When the underlying distributions were uniform and bimodal, the distributions were aggregated and applied equally across all conditions; colour had negligible effects on distinguishing between the distributions. When one distribution was bimodal and the other trimodal, observers' biases were compatible with learning some features of the underlying distributions, although they again failed to use this knowledge in the correct settings (Gekas et al., 2013). However, adults demonstrated intact ability to learn two distinct distributions at once in a distance reproduction experiment where the distributions were identified by symbolic cues (Petzschner et al., 2012). A major difference between the two studies is the fact that to distinguish between the distributions, Petzschner et al. (2012) used symbolic cues in the form of written instructions indicating whether the distance to be reproduced would be 'short' or 'long'; the cue distinguishing the distributions was explicit as opposed to implicit (colour) in Gekas et al. (2013). Arguably, even though people have been found to learn relationships between arbitrary paired cues over multiple sessions (Ernst, 2007), there is only weak evidence of colour as an effective cue for categorisation (Howard et al., 2013; Seydell, 2010). There have been a few exceptions to this; for example, where the stimuli correlate naturally with colour, as in the case of lighting (Kerrigan & Adams, 2013).

On the other hand, Bejjanki et al. (2016) showed that human participants can learn two prior distributions at once, with the distinct distributions identified by different stimulus locations and colours. Subjects had to locate an invisible 'bucket', which sometimes appeared on the right, and sometimes on the left side of the screen. On each trial, subjects would see some locations people have guessed the bucket to be in the past; these guessed locations appeared as either all green or white dots, depending on the distribution from which the bucket's true location was drawn. This study shows that participants could learn the two distinct distributions, suggesting that they can use colour information to differentiate between the distributions. Perhaps more importantly, Bejjanki et al. (2016) and Petzschner et al. (2012) used distributions of the same type (e.g., uniform distributions with different means, Petzschner et al., 2012), as opposed to complex multimodal distributions, such as a mixture of bimodal and uniform distributions as done by Gekas et al. (2013). It is therefore possible that using a mixture of distributions of different types might have been the limiting factor in some studies (e.g., Gekas et al., 2013).

Adult observers can also learn different prior distributions when presented sequentially. For instance, Vilares et al. (2012) changed the variance of the prior between blocks of trials and found that subjects' relative weights differed significantly between the prior conditions. A similar result was found in a study by Berniker et al. (2010): subjects relied more strongly on the current sensory input in the block with the wider prior variance. A caveat is that the strongest effects were observed when switching from a narrow to wide prior variance than the other way around. An enhanced sensitivity for narrower priors when experienced first has been reported by other studies, as well. For example, Miyazaki et al. (2005) demonstrated that subjects adjust their behaviour faster for narrow-to-wide changes than wide-to-narrow changes. Kording et al. (2004) also found that subjects decreased their reliance on the prior when its variance switched from narrow to wide; subjects did not change their weighting of the prior when it switched from wide to narrow. As was argued by Trommershauser et al. (2011), one possible reason for the slower adaptation to wide-to-narrow changes than the other way around is that when the change is from narrow to wide, observers experience a sudden surge of 'surprising' observations (i.e., unexpected under the narrow prior distribution). In contrast, when the change is from wide to narrow, observers have to notice that the stimuli are no longer appearing at the edges of the distribution; a process that may take a long time (Trommershauser et al., 2011).

Taken together, these studies suggest that learning and efficient use of prior information is heavily influenced by the complexity of the distribution (as was agued by Acerbi et al., 2014b): whilst the statistics of a normal distribution can be picked up after only ~200 trials (Berniker et al., 2010), adapting to a prior that has a bimodal distribution may take longer, requiring no less than thousands of trials (Körding & Wolpert, 2004). However, in none of these studies has it been possible to tease apart whether it is the learning of the prior or the efficient, statistical optimal use of priors that has been affected by complexity. In the next section, we mention briefly studies that have sought to tease these two apart.

1.5.2.1.3 Suboptimalities in the Learning or Use of Priors? Suboptimalities in perceptual decision-making are well documented; however, it is not known whether these suboptimalities arise due to difficulties in learning and representing the prior distribution or learning how to use it efficiently. For example, Acerbi et al. (2014b) eliminated the need to learn the prior by explicitly showing it on every trial and found that performance did not vary as a function of how complex the distribution was, with near identical performance across bimodal and simple Gaussian priors. The authors noted previous studies which contrast these data, showing that as compared with unimodal priors, evidence of efficient use of bimodal priors only begins to show when participants had performed thousands of trials (Körding & Wolpert, 2004) or when the data are pooled together (Chalk et al., 2010). Comparison between these studies led the authors to conclude that differences in behaviour with simple vs. complex distributions reported previously are likely driven by difficulty in learning more complex distributions, and not using them efficiently (Acerbi et al., 2014b).

This interpretation contrasts results by Bejjanki et al. (2016). They tested whether observers would be biased in the direction of the prior when no sensory information was available (i.e., on prior-only trials). Specifically, a ‘green’ or ‘white’ rectangle was rapidly flashed, indicating the side of the screen the ‘bucket’ was located. There was no significant difference between observers’ estimates and the ‘true’ prior means, suggesting that observers had learned the means of the underlying distributions. Also, the two priors were weighted differently throughout the experiment, suggesting that their variances were also tracked and taken into account. This suggests that the suboptimalities are likely to have arisen because of a failure to use the prior and sensory information efficiently, not because adult observers struggle with learning the statistics of distributions. However, these data cannot conclusively demonstrate this because the estimated prior variances were not measured directly. Future research is needed to determine the exact source of suboptimalities that have been previously observed.

1.5.2.2 Contextual Priors in the Developing Brain.

1.5.2.2.1 Infants Extract Statistical Regularities in Their Environments.

Infants can implicitly learn statistical regularities, and notice when these regularities had broken down, not only in their natural environments but also within experimental tasks themselves in the first year of life (Bulf et al., 2011; Fiser & Aslin, 2002; Kirkham et al., 2002, 2007; Tummeltshammer et al., 2017; Tummeltshammer & Kirkham, 2013; Wu et al., 2011). Saffran et al. (1996) presented a stream of syllables to 8-month-old infants and found that the 8-month-olds learned the transitional probabilities between syllables within minutes of exposure. Learning was measured through looking times where infants showed longer looking times to the novel syllable orderings compared to the familiar syllable orderings. Saffran et al.'s (1996) work has been interpreted as revealing a mechanism by which infants use conditional probabilities to segment spoken speech into syllables and words (Rohlf et al., 2017). Similarly, infants can learn statistical regularities in tone sequences around the same age (Saffran et al., 1999) and in visual patterns at 9 months (Fiser & Aslin, 2002). For example, Fiser and Aslin (2002) showed that 9-month-old infants who were presented with multi-element visual scenes preferred to look at the element pairs which appeared together more frequently than those who did not. Kirkham et al. (2002) presented 2-, 5- and 8-month-olds with repeated sequences of visual stimuli which followed a statistically predictable pattern. After, the infants were shown these same sequences but this time mixed with novel sequences which consisted of the same stimuli. Kirkham et al. (2002) found that infants in every age group looked longer at the novel sequences. This result suggests that infants can learn statistical regularities as early as 2 months of age.

1.5.2.2.2 Are Children as Good as Adults at Learning Statistical Regularities? Several studies have demonstrated that children are better at learning statistical relationships implicitly compared to adults (Janacsek & Nemeth, 2013; Nemeth et al., 2013). One example is the study of Jost et al. (2011), who sought to identify the neural correlates of statistical learning in 6- to 8-year-olds, 9- to 12-year-olds and adults: learning-related ERPs peaked earlier in children than adults, suggesting that children learned the statistical structure faster (Rohlf et al., 2020). Note however that not all studies have found that children are better at extracting statistical relationships than adults. For example, experiments by Saffran and colleagues (1996, 1999) found that 8-month-old infants are as good as, but no better than, adults in learning the statistical regularities embedded in auditory sequences. Better learning in older children and adults has also been reported (Kirkham et al., 2007; Maybery et al., 1995). Using a visual triplet learning task, Arciuli and Simpson (2011) showed that extracting statistical regularities improved between 5 and 12 years of age, and even further in young adults. This might be explained by the fact that learning stimulus triplets is more costly in terms of working memory and other cognitive resources than learning the conditional probability of two stimuli as used in many other studies (e.g., Emberson et al., 2015; Kouider et al., 2015). This interpretation is supported by studies showing that working memory, and executive functions more generally, develop over the course of childhood (Zelazo et al., 2008). As was pointed out by Rohlf et al. (2020), it is thus possible that greater cognitive costs (e.g., higher working memory demands) might have been the limiting factor in the study by Arciuli and Simpson (2011). Although the evidence is mixed, it seems as though more complex and a higher number of regularities (which draw more heavily on still-developing executive functions, Gur et al., 2012; Zelazo et al., 2008) seem to be better learned the older the children become.

1.5.2.2.3 Can Children Use Statistical Regularities to Improve Perception? Only a small number of studies have examined at what age children learn to incorporate prior knowledge optimally. Six-to-eight-year-olds, presented with noisy sensory information about the target stimulus (cloud of four dots at one of three levels of uncertainty) and explicit knowledge about previous stimuli (the prior), weighted the sensory information by its reliability but not the prior when estimating the location of a target (Chambers et al., 2018). Older children (> 9 years), on the other hand, weighted both sources of information (sensory information and prior) according to their reliabilities. These results were interpreted as showing that the basic mechanism allowing children to optimally integrate evidence from various sources does not develop before 9-10 years of age (Chambers et al., 2018).

An alternative explanation is that children's abilities to optimally combine are fundamentally limited by 'resource constraints', such as insufficient working memory capacity or cognitive control (Bejjanki et al., 2019; Bejjanki & Aslin, 2020; Lieder & Griffiths, 2019). This hypothesis was supported by Bejjanki et al. (2019) who investigated whether optimal integration of prior knowledge is influenced by task demands; when using a simple (Gaussian) prior they found adult-like rate of learning and use of prior information while with a bimodal prior they did not: instead, 6-7-year-old children systematically placed more weight on sensory signals than was optimal, even after a thousand trials. Working memory (and other executive functions, such as inhibitory control and flexibly switching) may be a factor here. Learning the mean and variance of a single distribution unlikely touches the limits of working memory – learning the statistics of two underlying distributions at the same time might. This idea – that children's ability to track and optimally use information about the statistics of the environment depends on the complexity of the environment, with greater complexity drawing more on cognitive resources – is consistent with resource rational accounts (Bejjanki & Aslin, 2020; Lieder & Griffiths, 2019). These accounts propose that there is an upper limit on the total amount of resource that can be invested; if this upper limit is much higher in adults, compared to younger children, then limitations in executive functions might be the constraining factor, and not that children are limited in, or ultimately lacking the ability to optimally combine.

1.5.2.2.3.1 Executive Function Continues to Develop. Executive functions are high-level cognitive processes, such as memory, sustaining attention and suppressing impulses (Diamond, 2013). We know that executive functions are linked to higher IQ (Arffa, 2007) and educational attainment (Best et al., 2011), and that they improve considerably between early childhood and mid-twenties (Best & Miller, 2010; Gur et al., 2012). Some of the strongest evidence in support of this comes from studies reporting increases in connectivity among fronto-parietal regions (Baum et al., 2017; Gu et al., 2015; Hagmann et al., 2010; Huang et al., 2015; Power et al., 2010), known to be important for executive function (Alvarez & Emory, 2006; Mansouri et al., 2017; Niendam et al., 2012; Satterthwaite et al., 2013). Altogether, multiple lines of evidence suggest that executive functions develop over the course of childhood and adolescence. If we assume that learning and using prior knowledge in a statistically optimal manner depends, in part, on the recruitment of executive functions, this could explain why children do not become adult-like in their abilities to optimally combine until later in childhood or adolescence. There are two executive functions– working memory and the ability to store multiple ‘rules and switch between them – that we believe are important for understanding how we learn and efficiently use novel statistics, in particular. In the following paragraphs, we explain what role working memory and cognitive flexibility might play in the ability to combine optimally.

1.5.2.2.3.1.1 The Importance of Remembering. Working memory is of special interest because it plays a central role in our ability to uncover statistical regularities in our environment (Baddeley, 2003, 2007; Bianco et al., 2020). As Bianco et al. (2020) argues, the detection of such regularities relies upon our memory storing individual experiences for long enough to allow us to figure out how they fit together. Working memory, therefore, is likely to play a crucial rule in accumulating information from the past, which in fact, is what we know as the prior. One could imagine that given that working memory continues to develop across adolescence and into early adulthood (Alloway et al., 2006; Crone et al., 2006; Gathercole et al., 2004; Luna et al., 2004, 2015; Thomason, Moriah et al., 2009), younger children would have a hard time remembering all past events, needed to infer the correct underlying regularities. Given these ties between prior learning and working memory and the late maturation of prefrontal regions, we might predict that adult-like integration of prior information is, at least in part, influenced by cognitive, and more specifically, working memory development.

1.5.2.2.3.1.2 The Importance of Remembering and Handling Multiple ‘Rules’. Secondly, it is possible that the ability to store and alternate between multiple ‘rules’ may play a role in learning to weight sensory and prior information according to their reliabilities, even if it is not required to support *learning* of the prior. Bejjanki et al.'s (2019) finding that 6-7-year-old children were adult-like in how they weighted sensory information when only sensory uncertainty varied, but now when both prior and sensory uncertainty varied (increased task complexity) supports this assertion. Similarly, it has been shown that 4-year-olds could store information and suppress an unwanted response at the time when task ‘rules’ remained stable (Davidson et al., 2006) but could not flexibly switch between task rules in similar ways to adults as late as 13 years of age (see also Luciana & Nelson, 1998). Such results indicate that the ability to store and flexibly switch between rules might play a role in learning to weight by reliability, especially in children, when this ability is still developing.

1.5.2.2.4 Filling the Gaps. Very few studies have explored the development of prior and sensory information (Bejjanki et al., 2019; Chambers et al., 2018), so we know very little about the age-related changes in learning and efficient use of novel statistics throughout childhood, or the factors that may underlie the development of these abilities. Because another aspect of efficient information integration, multisensory cue combination, also does not develop until the age of around 9 years (e.g., Gori et al., 2008; Nardini et al., 2008), one could imagine that they share a common, central problem with representing, or computing with probabilities. Another possibility – which is specific to learning and use of novel statistics – is that children are slower to accumulate noisy sensory information over time. This slowness could be explained by the fact that the processes of evidence accumulation and reliability-weighting depend, at least in part, on executive functions (e.g., working memory) which are still very much in development (Best & Miller, 2010; Gur et al., 2012). Yet, this hypothesis has not yet been tested. To fill this gap, in Chapter 3, we tracked the detailed time course of adaptation to novel statistical regularities (Experiment 1) and probed two potential sources of slower tuning of children's behaviour to task statistics: i) limited working memory for storing the prior (Experiment 2) and ii) limited executive function abilities for learning and implementing multiple weighting rules simultaneously (Experiment 3).

1.5.2.3 Individual Differences. Healthy adult observers show considerable individual differences in how well they can detect and efficiently use statistical regularities in their environments (Frost et al., 2015; Misyak et al., 2010; Siegelman & Frost, 2015). As we discussed in Section 1.5.2.2.3.1.1 above, working memory is critical to our ability to store recent experiences in memory for long enough to detect any underlying patterns. We know that working memory capacity varies greatly between individuals (Kane et al., 2007). It has been shown that it is advantageous to have a higher working memory capacity when making decisions under uncertainty (Bagneux et al., 2013; Cui et al., 2015), as well as retaining and implementing a new decision or rule (Duncan et al., 2012; Pereg & Meiran, 2019).

1.5.2.3.1 Filling the Gaps. As elaborated above, there are inter-individual differences in the ability to detect statistical patterns. One source of this variance could be individual differences in working memory capacity. Yet, this hypothesis has not yet been tested. To fill this gap, in Chapter 4, we asked whether working memory capacity, in particular visuospatial working memory capacity, could uniquely explain a proportion of the variance in how close to optimal adults are.

1.6 The Caveat

As many researchers have pointed out (Ma, 2012; Ma & Jazayeri, 2014; Maloney & Mamassian, 2009), a key limitation of psychophysical studies is that there is no definite evidence that observers are in fact performing Bayesian inference, rather than a simpler heuristic, such as reinforcement learning, which approximates it. Laquitaine and Gardner (2018) showed that a Switching observer that alternates between prior mean and sensory evidence can capture experimental data from a motion direction estimation task better than a Bayes-optimal observer, despite the mean and standard deviation of estimates matching well with optimal predictions. Norton et al. (2019) also found that their data was better accounted for by an exponential averaging model with a bias towards equal priors, which was, nonetheless, indistinguishable from a flexible variant of the Bayesian model with a bias towards equal probability. These studies suggest that just because a Bayesian model can explain broad qualitative features of the data, it does not mean that subjects perform Bayesian inference. It is important to keep this caveat in mind when interpreting our and others' findings. One way to check whether observers perform some form of Bayesian inference is to look at whether subjects can instantly transfer probabilistic information from one context to the other (Beierholm et al., 2009a; Maloney & Mamassian, 2009). It has previously been shown that likelihood variance learning can immediately transfer to a new prior (Sato & Kording, 2014) and that priors whose distribution changed on each trial could be immediately computed with

(Acerbi et al., 2014a). However, some studies found no evidence of transfer (Frost et al., 2015; Hewitson et al., 2018; Yin et al., 2019) and others suggested that the ability to transfer differs depending on whether the observer is transferring information about the mean or the variance of the distribution (Fernandes et al., 2014; Feulner et al., 2020).

1.6.1 Filling the Gaps

In Section 1.6, we showed that although there are many studies on transfer, we know very little about the exact circumstances in which observers are able to generalise what they learn in one situation to new situations. For example, it may be that generalisation becomes more difficult as the complexity of the environment increases. Thus, across several experiments in Chapter 2, observers localised a hidden target from visual signals (likelihood) and previous target locations (prior). Crucially, we manipulated the variance of the likelihood and the prior. Another important aspect of the experiments reported in Chapter 2 was the interleaved (as opposed to block-based) design for changes in these parameters, which had several consequences. For example, it is likely that compared to some previous studies of transfer (e.g., Sato & Kording, 2014), our task was substantially more difficult since the prior and likelihood variances changed unpredictably trial to trial (see also Bejjanki et al., 2016). This also prevented subjects from using a template-match strategy. Crucially, we introduced a new level of likelihood variance in the last two of the five experimental blocks, which allowed us to test the prediction of transfer: that the change in weights in response to the new likelihood variance would be immediate. The key research question of the experiments in Chapter 2 was: did adult observers *immediately* change their weight on the sensory cue, or did they need more time to adapt to the new variance?

1.7 Conclusions

The work in this thesis aims to better understand how people learn and make efficient, statistically optimal use of novel statistical distributions. In Chapter 2, we examined one potential limitation to using statistics in an optimal, that is Bayesian, way in adults – environmental complexity. In Chapter 3, we traced the time course of adaptation to such statistics in children and adults and probed some factors that may have affected age differences in the extent and rate of this adaptation. Lastly, in Chapter 4, we delved into individual differences in working memory in adults and asked whether these predict the extent to which different individuals efficiently weighted sensory and prior information. Altogether, these studies speak to questions about the limits and the development of abilities to use novel environmental statistics in an efficient, statistically optimal manner.

1.8 Experimental Chapters

Chapter 2: Bayesian transfer in a complex spatial localization task

Whether the brain performs some form of Bayesian inference (Chater et al., 2006; Dayan et al., 1995), or just approximates it, is open for debate. There are a number of different ways to rule out simple heuristics (e.g., reinforcement learning) but probably the simplest one is to test whether observers generalise previously learned information to novel situations (‘Bayesian transfer’; Maloney & Mamassian, 2009). Across three experiments in **Chapter 2**, therefore, we apply the ‘Bayesian transfer’ approach to investigate whether observers combine information from past and current signals in a Bayesian manner in a complex task, or whether they implement a heuristic version of the optimal strategy by forming a look-up table. We were also interested in whether their strategy might change if **(1)** they were given more information about the underlying statistics (Experiment 2), **(2)** we increased the exposure time to the different prior and likelihood combinations (Experiment 3). We investigated this in adult participants only.

Chapter 3: Development of efficient adaptation to novel task statistics

Research exploring whether and at what age children learn to integrate prior knowledge is limited, as is our understanding of either overall age changes in learning and efficient use of novel statistics in childhood, or the factors that might underlie development of these abilities. Moreover, most previous studies averaged performance over the experiment as a whole (Bejjanki et al., 2019; Chambers et al., 2018). We therefore do not know what the time course of adapting to novel, changing task statistics over time is, and whether it differs between adults and children. Studying the time course is important because it can show whether children are generally incapable of using prior statistics efficiently, or whether it takes them longer to learn how to make efficient use of such statistics. Therefore, in **Chapter 3**, we traced for the first time the time course of how children and adults adapt. In two follow-up experiments we probe two potential sources of slower tuning of children's behaviour to task statistics: i) limited working memory for storing the prior (Experiment 2) and ii) limited executive function abilities for learning and implementing multiple weighting rules simultaneously (Experiment 3). We investigated this in adults and children aged between 6 and 11 years.

Chapter 4: Individual differences in working memory capacity and prior weighting

We know that there is considerable individual variation in the ability to detect and extract statistical regularities even in healthy adults (Frost et al., 2015; Misyak et al., 2010; Siegelman & Frost, 2015). The storage capacity of working memory, which is also known to differ between individuals (Kane et al., 2007), will likely play a key role in the acquisition of statistical regularities. Therefore, in **Chapter 4**, we sought to test whether working memory capacity can predict inter-individual variation in prior weighting. We investigated this in adults using an individual differences approach.

Chapter 2

Bayesian transfer in a complex spatial localization task

2.1 Introduction

Imagine that you are trying to give your cat a bath, but as soon as it sees the bathtub, it gets scared and runs away to the garden (Kording et al., 2007; Vilares & Kording, 2011). So, you are walking around your garden, trying to figure out where your cat has hidden, and you hear a "meow" *(auditory cue)*. This perceptual cue is useful but not perfectly reliable and will not allow you to pinpoint exactly the cat's position. However, from previous experience, you may have learnt that your cat often hides in the bushes, furthest from the pond *(priors)*. The uncertainty in the two pieces of information that you have (the auditory cue and the prior information) allow them to be expressed as probability distributions over location and the optimal strategy for estimating the cat's location is to integrate the sensory and prior information according to the rules of Bayesian Decision Theory (BDT). Recent studies show that people behave as if they deal with uncertainty in this way, for example, when estimating the position of a hidden target (Berniker et al., 2010; Körding & Wolpert, 2004; Tassinari et al., 2006; Vilares et al., 2012), direction of motion (Chalk et al., 2010), speed (Stocker & Simoncelli, 2006; Weiss et al., 2002), or the duration of a time interval (Acerbi et al., 2012; Ahrens & Sahani, 2011; Jazayeri & Shadlen, 2010; Miyazaki et al., 2005). In all of these studies, human observers integrated knowledge of the statistical structure of the experiment (acquired from feedback in previous trials) with sensory information, taking a weighted average according to their relative reliabilities in order to maximise his or her score on the task (Ma, 2012). However, other studies report sub-optimal behaviour, finding that even though observers take into account the uncertainty of the current and prior information, the weights do not match those of an ideal Bayesian observer (Bowers & Davis, 2012; Jones & Love, 2011; Rahnev & Denison, 2018). The fact that human performance ranges from close-to-optimal to largely suboptimal suggests that Bayesian models may describe behaviour well in some cases, but not in others. Understanding when BDT can and cannot provide an accurate model of human behaviour is an important step towards understanding the computations and approximations used by the brain to support adaptive behaviour.

One factor that may influence how close performance is to BDT (“optimal”) predictions is task complexity. For example, Bejjanki, Knill, and Aslin (2016) asked observers to estimate the position of an unknown target (a bucket at a fairground), whose true location was randomly drawn from one of two Gaussian distributions, with different means and variances (*priors)*. On each trial, eight dots were drawn from a Gaussian distribution centered on the true location with either a low, medium or high variance to form a dot cloud that served as a noisy cue to target location (*likelihoods)*. To successfully estimate the position of the target, subjects could use both the likelihood, obtained from the displayed dots, and the prior, obtained from the distribution of previous target positions that they could learn from the trial-by-trial feedback. The study found that subjects adjusted their responses according to the reliability of sensory and prior information, giving more weight to the centroid of the dot cloud (likelihood) when the variance of the prior was high and the variance of the likelihood was low; a signature of probabilistic inference (Ma, 2012). More generally, these results are also in agreement with previously described work, which used *a single prior change* (e.g., Berniker et al., 2010; Vilares et al., 2012). However, unlike in the studies using only a single prior change, the weight placed on the likelihood differed from that of the ideal-Bayes observer whenever the likelihood uncertainty was medium or high. The magnitude of the difference varied with both likelihood and prior uncertainty. As the likelihood became more uncertain, the difference from optimal increased, participants placing more weight on the likelihood than optimal. In addition, the difference from optimal was greater in the narrow prior condition. In a follow-up experiment, participants experienced one prior distribution only, with double the amount of trials used in the original study, finding that subjects’ weights on the likelihoods approach optimal with increasing task exposure, suggesting more time is required to accurately learn the variances of the prior distributions and that learning is disrupted when trying to learn two distributions simultaneously (Bejjanki et al., 2016).

Even in cases when likelihood-weighting might match the prediction of an ideal-Bayes observer, Maloney and Mamassian (2009) noted that such “optimal” or “near-optimal” performance alone is not enough to show that the brain is following

Bayesian principles. Maloney and Mamassian (2009) showed that a reinforcement-learning model that learns a mapping from stimulus to response (learning a separate look-up table for each prior and likelihood pairing in the types of tasks discussed here) will also appear to optimally weight prior and likelihood information without learning the individual probability distributions.

Maloney and Mamassian (2009) suggested that the two models may be distinguished by asking whether subjects are able to immediately transfer probabilistic information from one condition to another (hereafter known as *Bayesian transfer*). These transfer criteria are a strong test for use of Bayesian principles because they make very different predictions for how the observer will behave when presented with a new level of sensory noise halfway through the task. If people follow Bayesian principles, we would expect them to *immediately* adapt to the new sensory uncertainty, and integrate it with an already-learnt prior, without any need for feedback-driven learning. On the other hand, the reinforcement-learner would require a certain amount of exposure to the new likelihood and prior pairing (with feedback) in order to form a look-up table that could then lead to optimal performance.

To our knowledge, only one study has tested Bayesian transfer in the context of sensorimotor learning. Sato and Kording (2014) tested the ability of participants to generalise a newly learnt prior to a previously learnt likelihood. In their task, Sato and Kording (2014) first trained participants to complete the task when only a single Gaussian prior was present (either narrow or wide) that could be paired with either a low or high uncertainty likelihood by giving feedback on every trial. After 400 trials, the prior switched to the other level of uncertainty (narrow to wide or wide to narrow) and for the following 200 trials, participants saw the new prior paired only with one of the likelihoods (either low or high) and continued to receive feedback. In the second part of the experiment, subjects still saw the second prior variability, but now with the first likelihood again, which they had so far only seen paired with the first prior. They did not provide any feedback in this part of the task to examine how subjects transferred their knowledge of the prior to the new likelihood. The weight placed on the likelihood in the newly-reintroduced likelihood condition was immediately different to the weight placed on the same likelihood before the change in the prior. In other words, participants' behaviour in this likelihood condition changed dependent upon prior uncertainty without any explicit training with this prior and likelihood pairing. This is evidence of Bayesian transfer and hence, that participants solved the task using Bayesian principles – representing probabilities – rather than a simpler strategy such as a look-up table learned by reinforcement.

Whether the same will hold in more complex scenarios is unclear. Indeed, it has been repeatedly pointed out that exact Bayesian computations demand considerable computational resources (e.g., working memory, attention) such that the brain might not be able to perform these computations in more complex tasks and will instead resort to heuristics (Beck, Ma, Pitkow, Latham, & Pouget, 2012; Gardner, 2019). We have already seen that performance in more complex tasks is far from optimal (e.g., Bejjanki et al., 2016), suggesting that there are limits to humans' ability to learn and optimally integrate prior distributions with sensory information when tasks become more complex. Establishing the limits of BDT as a model of human behaviour will inform models of information processing in the human brain.

Here we ask whether people will show Bayesian transfer in a complex situation with multiple priors and likelihood variances, similar to Bejjanki et al. (2016). We report three experiments in which a target is sampled from one of two possible prior distributions (with different means and different variances) and then cued with one of three possible likelihood variances (with the variance itself also displayed). Likelihood and prior variances were identical to those used in Bejjanki et al. (2016) in terms of visual angle, in order to match the true (objective) reliabilities of the cue and prior across the studies. The first two experiments tested for Bayesian transfer by only introducing the high likelihood variance in blocks 4 and 5 of the task. The only difference was that in the second experiment, participants were explicitly told that the prior variances differed, to test whether this would promote closer-to-optimal performance. The last experiment was used to check whether removing the additional burden of transfer allows participants to learn the complex environment correctly by presenting all likelihood conditions from the start of the experiment – a replication of Bejjanki et al. (2016).

To summarise, in the first experiment, we found that observers did not show evidence for Bayesian transfer. When a new high variance likelihood was introduced in blocks 4 and 5, they did not weight it less than the familiar medium variance likelihood. This is at odds with the idea that observers perform full Bayesian inference, combining prior and likelihood based on relative variances, and thus does not provide evidence for Bayesian transfer. In the second experiment, observers did show evidence for transfer, weighting the newly-introduced high variance likelihood significantly less than the medium variance likelihood. However, in both experiments, the weights placed on the medium and high variance likelihoods were much higher than optimal. These weights remained sub-optimal in the final experiment where all likelihood variances were present from the start of the task. These results extend our knowledge of how potentially-Bayesian perceptual processes function in complex environments.

2.2 Experiment 1: Testing Transfer to a New Level of Likelihood Variance

In the first experiment, we tested whether Bayesian transfer would occur in a complex environment with two priors, similar to the one used by Bejjanki et al. (2016). We trained participants on a spatial localization task with two likelihood variances and two prior distributions (with different means and variances). In initial training, they were exposed to all four combinations (trials interleaved), with feedback. If, like participants in Bejjanki et al. (2016), they weighted the likelihood and the prior differently across conditions in line with their differing reliabilities, this would show that they had learned and were using the priors. However, such reliability-weighting could either be done via Bayesian inference – representing probabilities – or via a simpler strategy akin to learning a look-up table (Maloney & Mamassian, 2009). To distinguish these possibilities, after the training trials, we tested for “Bayesian transfer” by adding a new higher-variance likelihood distribution to the task. If participants deal with this newly-introduced likelihood in a Bayesian manner, they should immediately rely less on this new likelihood information than they did on the likelihoods in previously-trained conditions. Alternatively, if their initial learning is more rote in nature (i.e., more like a look-up table), participants would begin to place a different weight on the new likelihood only after extensive training with feedback.

2.2.1 Methods

2.2.1.1 Overview. Subjects performed a sensorimotor decision-making task on a computer monitor where they estimated the horizontal location of a hidden octopus. The true location was sampled from one of two distinct Gaussian distributions that differed in mean and variance (narrow or wide priors). On each trial, the relevant prior distribution was indicated by the colour of the likelihood information - eight dots that were described to the participant as the "tentacles" of the octopus. The horizontal locations of the eight dots were sampled from a Gaussian distribution centred on the true location that had either low, medium, or high variability (the likelihood). To estimate the octopus' position, participants could use (although this was never explicitly mentioned) both the likelihood and prior information, with the subjects able to learn the latter via trial-to-trial feedback. Participants completed five blocks of trials. Crucially, in blocks one to three only the low and medium likelihood variances were paired with the narrow or wide priors. The high likelihood condition was only introduced in blocks four and five to test for evidence of Bayesian transfer.

2.2.1.2 Participants. Participants were recruited from the Durham Psychology department participant pool, Durham University newsletter, and by word-of-mouth. Twenty-six participants were recruited in total (13 female, mean age: 20.1, age range: 18-30 years). All participants had normal or corrected-to-normal visual acuity, and no history of neurological or developmental disorders. Each participant received either course credits or a cash payment of £10 for their time.

2.2.1.3. Ethics. Ethical approval was received from the Durham University Psychology Department Ethics Board. All participants gave written, informed consent prior to taking part in the study.

2.2.1.4. Stimuli and Apparatus. Stimuli were displayed on a 22-inch iiyama monitor (1680 x 1050 pixels), viewed at a distance of 60cm, using the Psychophysics Toolbox for MATLAB (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). The stimuli were set against a blue background (to represent the sea).

The position of the octopus was sampled from one of two Gaussian distributions (priors): the narrow (standard deviation of $\sigma_{pl} =$ 1% of screen width) or wide (standard deviation of $\sigma_{ph} =$ 2.5% of screen width) priors. The side of the screen associated with each prior was counterbalanced across participants. One was always 35% of the way across the screen (from left to right), and the other 70%. When the narrow prior was centered on 35%, for example, the wide prior had a mean in the opposite side of the screen (i.e., to the right, centered on 70%). When the octopus appeared on the left-hand side (drawn from the prior centered on 35%) it was white, and when it appeared on the right (drawn from the prior centered on 70%) it was black.

At the beginning of each trial, a cloud of eight dots (0.5% of screen width in diameter) appeared on the screen. The horizontal position of each dot was drawn from a Gaussian distribution centered on the true octopus location with either low ($\sigma_{ll} = 0.6\%$ screen width), medium ($\sigma_{lm} = 3\%$ screen width), or high ($\sigma_{lh} = 6\%$ screen width) standard deviation (in the following referring to as low, medium and high variance likelihood conditions). The horizontal positions of the dots were scaled so that their standard deviation (SD) was equal to the true SD ($\sigma_{ll},\ \sigma_{lm}\ or\ \sigma_{lh}$) on each trial while preserving the mean of the dots. We performed this correction so that participants would "see" the same variability across trials for each likelihood condition. This ensures that an observer who computes the reliability for the likelihood information trial by trial would always calculate the same value within likelihood trial types. The vertical positions of the dots were spaced at equal intervals from the vertical center of the screen, with half of the dots appearing above, and the other half, below the center. The vertical distance between each dot was fixed and equal to 1% screen width). Given that the vertical positions of the dots were fixed, only the horizontal position of the target was relevant. Participants estimated location only along the horizontal axis by moving a vertical green rectangle (measuring 1% of screen width in width and 3% of screen width in height) left or right, making this a one-dimensional estimation task. Participants received feedback in the form of a red dot (0.5% of screen width in diameter) that represented true target position.

The combination of two priors and three likelihoods led to six trial types (all possible prior x likelihood pairings). The task was split into five blocks of trials with

300 trials per block. In the first three blocks of the task only four trial types were used (75 trials of each paring), with the high likelihood condition not shown in combination with either prior. The high likelihood condition was introduced in blocks four and five (50 trials per paring), in order to test for Bayesian transfer. Within each block, all trial types were interleaved. The trials were broken into runs of 20 trials. Within each run the trials for each prior type were arranged such that an ideal learner would have an exact estimate of the mean and variance of the prior distributions if evidence was accumulated over those 20 trials.

We also included prior-only trials where subjects were told that a black/white octopus was hiding somewhere on the screen and instructed to find it. No sensory information was provided (no likelihood information). These trials were interleaved with the rest of the trials (one every 9 trials for each prior), and there were 83 trials in total, for each of the priors (narrow and wide).

2.2.1.5. Procedure. Participants were instructed to estimate the position of a “hidden” octopus, indicating their estimate by adjusting the horizontal location of a “net” (green rectangle). Each trial started with the presentation of eight dots that remained on screen until the end of the trial (the likelihood information, described to the participants as the tentacles of the octopus) (Figure 2.1A). The eight dots could have one of three levels of uncertainty: low, medium, or high variance likelihood trials (Figure 2.1B). When the level of uncertainty was higher, the dots were more dispersed on the screen and, therefore, were a less reliable indicator of the true location of the octopus. Participants used the mouse to move the net to their guessed position, using a right click to confirm their choice (no time limit). Following a response, the true position of the octopus was shown as a red dot on the screen. Over the course of the experiment, the feedback served as a second cue to location since the true locations of the black and white octopi were drawn from different distributions. In other words, participants could learn a prior over each octopus’ location. We provided performance feedback on a trial-to-trial basis so that the priors could be learned. Specifically, subjects could potentially learn that the two sets of octopi (black/white) were drawn from separate Gaussian distributions centred at different locations on the screen and with differing levels of uncertainty (narrow and wide variance prior trials).

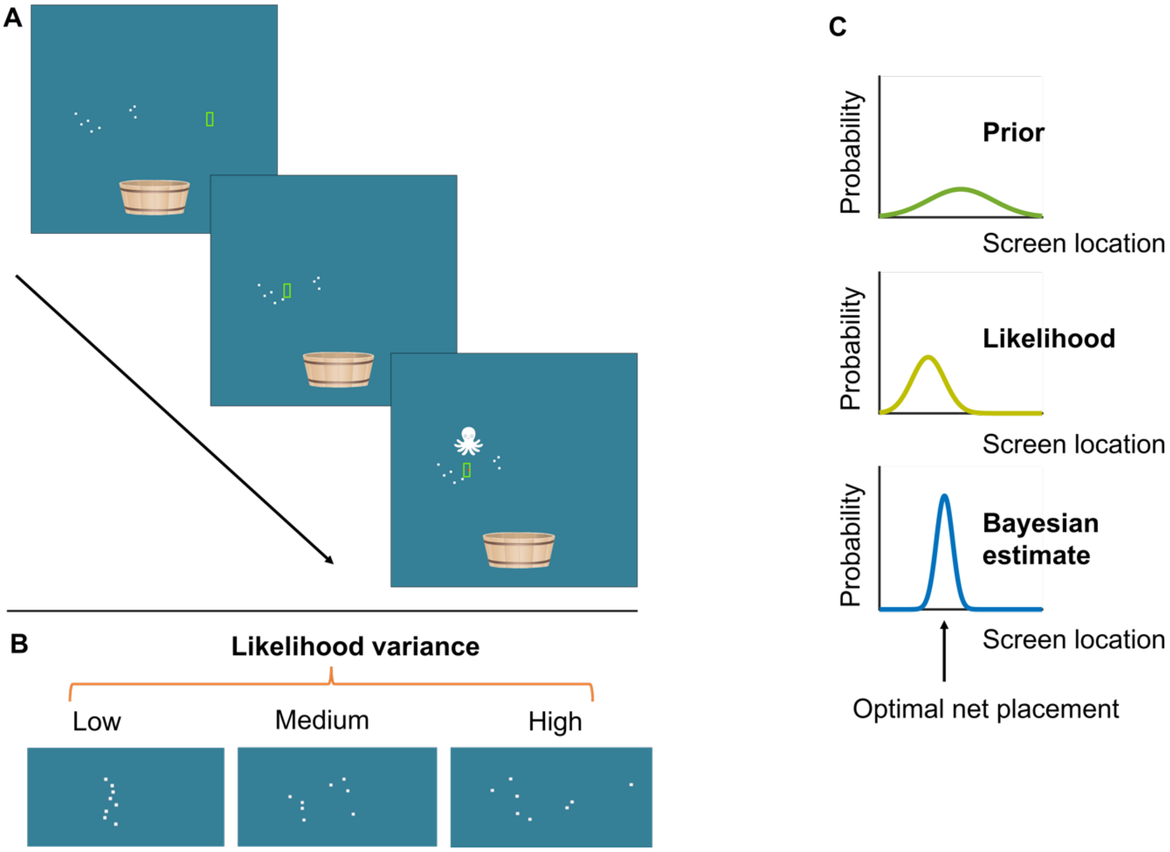

Figure 2.1. A) Illustration of the task. Participants were asked to estimate the position of a hidden target (the "octopus", represented as the red dot) by horizontally moving a net (green vertical bar). At the beginning of each trial, participants were given noisy information about the location of the hidden target in the form of eight dots (the likelihood). Participants then moved the net to the estimated location and clicked to confirm their response, after which the actual target location was displayed. If the target was inside, or overlapped with, the net, a score was added to the participant's score. B) The three likelihood variances. C) Illustration of a Bayes-optimal observer. A Bayesian observer would combine information about the prior uncertainty (learnt from the distribution of previous target locations) with the likelihood information on a given trial to optimally estimate the target location.

To keep participants engaged, we incorporated an animation when the participant picked the right location of a cartoon octopus moving into a bucket centred at the bottom of the screen. In addition, participants would get one point added to their score if they “caught” the octopus. An octopus was caught if the true octopus position overlapped with the net placement by at least 50% of the red feedback dot’s size. The cumulative score was displayed at the end of each trial. Participants completed 5 blocks of 300 trials each for a total of 1500 trials. The five experimental blocks were performed in succession with a short break between each one. The experiment duration was approximately an hour and a half.

2.2.1.6. Data Analysis. For each individual participant, we regressed estimated octopus position against the centroid (mean) of the cloud of dots (likelihood) on each trial. All regression analyses were done using a least squares procedure (the polyfit function in MATLAB). The slope of the fitted regression line quantifies the extent to which participants rely on the current sensory evidence (likelihood), as opposed to prior information. A slope of one suggests that participants only use likelihood information and a slope of zero suggests that participants rely only on their prior knowledge, ignoring the likelihood. A slope between zero and one suggests that both likelihood and prior information are taken into account, and the steeper the slope, the more participants rely on the likelihood and less on the prior information. Accordingly, we will refer to the fitted slope values as the weight placed on the likelihood.

We also computed the weight that would be given to the likelihood in each condition by an ideal Bayesian observer with perfect knowledge of the prior and likelihood distributions (see Figure 2.1C for an illustration). The optimal weight on the likelihood was computed as:

$$w_{optimal} = \frac{\frac{1}{\sigma_l^2/n}}{\frac{1}{\sigma_l^2/n} + \frac{1}{\sigma_p^2}}$$

where σ_l^2 is the variance of the likelihood, n is the number of dots that indicate the likelihood (in this case, there were 8 dots), and σ_p^2 is the variance of the prior.

To determine the proportion of the variance in responses that is accounted for by change in the estimate from the sensory cue, the coefficient of determination (R^2) was calculated by linearly regressing participants' responses against each estimate participants could have taken from the cue (i.e., arithmetic mean, robust average, median or mid-range). This was done for the combined data of all subjects in each experiment, across all blocks and trial types (prior and likelihood pairings). The estimate with the highest R^2 value was taken to be the estimate participants had most likely used.

Statistical differences were analysed using repeated-measures ANOVA with a Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) of the degrees of freedom in order to correct for violations of the sphericity assumption if $\epsilon \leq 0.75$ and a Huynh-Feldt correction otherwise.

We discarded a trial from analysis if the absolute error for that trial was in the top 1% of all absolute errors, computed separately for each prior and likelihood pairing across all blocks and participants (this rule excluded at most 13 trials per pairing for an individual subject).

2.2.2 Results and Discussion for Experiment 1

We first checked whether subjects took the mean as an estimate from the sensory cue, and not a heuristic, such as the robust average. In tasks similar to ours (Bejjanki et al., 2016; Chambers et al., 2018; Vilares et al., 2012), authors assume that observers use the mean of the dots as their best estimate of true location from the likelihood information. However, we did not explicitly tell our participants how the eight dots that formed the likelihood were generated, or that the best estimate they could take from them was their mean position, leaving open the possibility that observes may have taken a different estimate from the cue than the mean (de Gardelle & Summerfield, 2011; Van Den Berg & Ma, 2012). The mean horizontal position was found to explain the most amount of variance in participants' responses ($R^2 = 0.996$), relative to the robust average ($R^2 = 0.995$), median ($R^2 = 0.995$) or the mid-range of the dots ($R^2 = 0.992$). This suggests that the mean of the dots is the estimate that participants take from the sensory cue.

We then examined whether the weight participants placed on the likelihood, relative to the prior, varied with respect to trial type (prior/likelihood pairing) for all the trial types present from the beginning of the experiment. Without this basic result – a replication of the pattern found by Bejjanki et al. (2016) – we could not expect them to transfer knowledge of the learnt prior distributions to the new high variance likelihood in the later blocks. This was a qualified success: for these trial types (blue and green bars in Figure 2.2), participants showed the predicted pattern, but placed more weight on the likelihood than is optimal (compare bar heights to dashed lines in Figure 2.2), in line with previous research (Bejjanki et al., 2016; Tassinari et al., 2006). We conducted a 2 (narrow versus wide variance prior) x 2 (low versus medium variance likelihood) x 5 (block) repeated measures ANOVA with the weight given to the likelihood (the displayed dots) as the dependent variable. These results are shown in Table 2.1 and summarised here. There was a main effect of prior variance, with less weight on the likelihood when the prior was narrower ($p < .001$). There was also a main effect of likelihood variance ($p < .001$), where participants relied less on the medium variance likelihood. However, there was also a significant interaction effect of likelihood and prior ($p = .001$). When the prior was narrow, the decrease in reliance on the likelihood was smaller as the likelihood variance increased ($t(25) = 3.57, p = .001$).

We found a main effect of block ($p < .001$) and an interaction between block and likelihood ($p = .014$), with the medium variance likelihood weighted significantly differently across blocks (simple main effect of block, $F(4,100) = 5.84, p < .001, \eta^2_{partial} = .189$, weights decrease with increasing exposure), but not the low variance likelihood (no simple main effect of block, $F(4,100) = 1.64, p = .169, \eta^2_{partial} = .063$). This suggests that participants adjusted, through practice, their weights on the medium variance likelihood, getting closer to optimal.

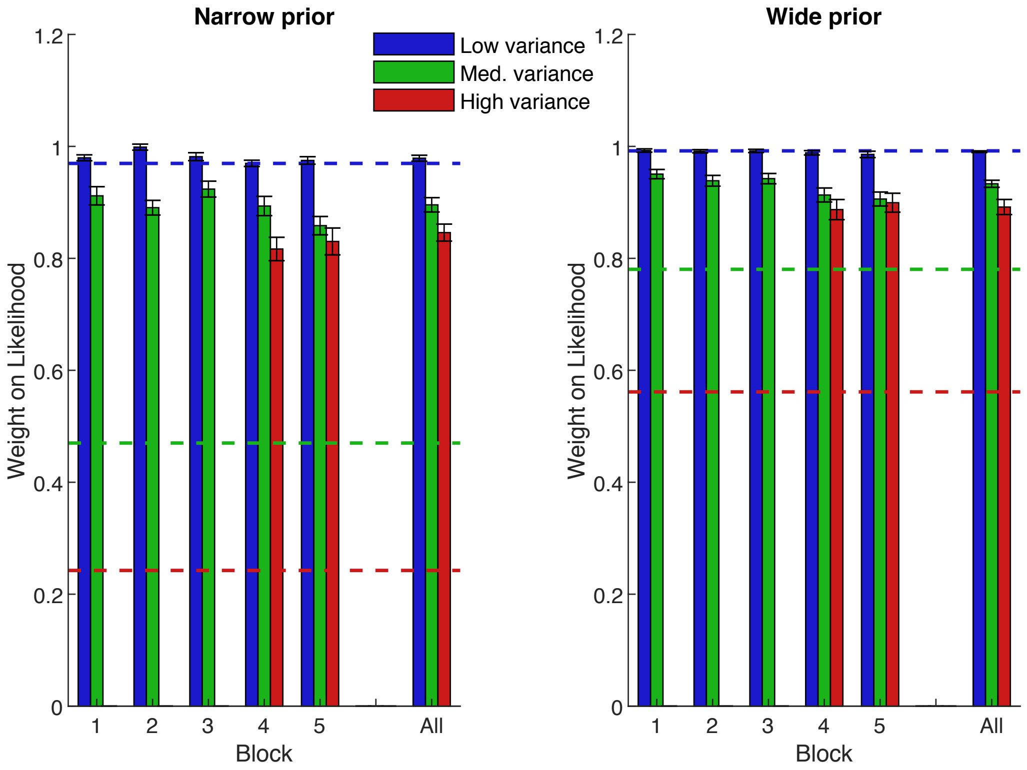

Figure 2.2. Mean weight placed on the likelihood information, separated by block and prior width in Experiment 1. Lower values represent a greater weight on the prior. Blue is low-variance likelihood (a tight array of dots), green is medium-variance likelihood (dots somewhat spread out), red is the later-introduced high-variance likelihood (highly spread out dots). Dashed lines show optimal-predicted values. Error bars are +/- 1 SEM. The far right is the average over blocks.

Table 2.1

Results from a 2 (prior) x 2 (likelihood) x 5 (block) Repeated Measures ANOVA for the likelihood variances present from the beginning of the task (low and medium) in Experiment 1

	F	*df, df_{error}*	*p*	Effect size (η_p^2)
Likelihood	104.40	1, 25	<.001	.81
Prior	28.08	1, 25	<.001	.53
Block	7.72	4, 100	<.001	.24
Likelihood x prior	12.77	1, 25	.001	.34
Likelihood x block	3.32	4, 100	.01	.12
Prior x block	.51	4, 100	.21	.06

Examination of the prior-only trials shows successful learning of the priors. On average, subjects' responses were not significantly different from the prior mean in the wide prior condition ($t(25) = -0.77, p = .450$). They were significantly different in the narrow prior condition ($t(25) = -2.78, p = .010$), although the bias was extremely small (95% CI: [0.06,0.41] percent of the screen width to the left). The median standard deviation of responses for all subjects was 1.4% (narrow prior) and 2.5% (wide prior): almost identical to the true prior SDs of 1.3% and 2.5%, respectively.

Participants qualitatively followed the predicted optimal pattern of reweighting: like the dashed lines (predictions) in Figure 2.2, actual likelihood weights (bars) were higher for the wide prior (right) than the narrow prior (left), and higher for the low variance likelihood (blue) than the medium variance likelihood (green). However, comparing bar heights with dashed lines (predictions) shows that quantitatively, their weights were far from optimal. Participants systematically gave much more weight than is optimal to the likelihood when its variance was medium (Figure 2.2, green bars vs lines – $p < .001$ in all blocks for the medium likelihood when paired with either prior). This over-reliance on the likelihood is in line with previous studies (e.g., Bejjanki et al., 2016), although stronger in the present study. Participants, therefore, accounted for changes in the probabilities involved in the task (e.g., weighted the likelihood less when it was more variable), but did not perform exactly as predicted by the optimal strategy.

Having found that participants' performance was in line with the predicted patterns, we could then ask if they would generalise their knowledge to the new high likelihood trials added in blocks 4-5 ("Bayesian transfer"), as predicted for an observer following Bayesian principles. This should lead immediately to a lower weight for the new high variance likelihood than the familiar medium variance likelihood. By contrast, lack of a significant difference in weights between the medium and high likelihood trial types would suggest that the observer is employing an alternative strategy, such as simply learning a look-up table. To test this, we performed a 2 (prior) x 3 (likelihood) x 2 (block) repeated measures ANOVA (summarise only blocks 4 and 5 – those with all likelihoods present). These results are shown in Table 2.2 and summarised here. As above, we found a main effect of likelihood, participants placing less weight on the likelihood as it became more uncertain ($p < .001$). However, post-hoc analyses showed that the weight placed on the high likelihood was not significantly lower than the weight placed on the medium likelihood ($p = .103$). Only the comparison of the weights placed on the likelihood in low and high variance trial types was significant ($p < .001$). Moreover, there was no main effect of block ($p < .28$), nor an interaction between block and likelihood ($p = .48$), suggesting that the weight placed on the newly introduced likelihood variance did not decrease with increasing exposure.

Finally, we compared mean weights in block 5 against the optimal Bayesian values for each prior and likelihood pairing. In the low variance likelihood trials we did not observe significant deviation from the Bayesian prediction, irrespective of prior (low variance likelihood, narrow prior: $t(25) = .784, p = .440$); low variance likelihood, wide prior: $t(25) = -1.12, p = .270$). Subjects' weights differed significantly from optimal in all other conditions ($p < .001$ in all cases).

Overall, our results do not exactly match the predictions of a Bayesian observer because we find only weak evidence of Bayesian transfer. Specifically, while we find a main effect of likelihood, the weight on the high variance likelihood is not significantly different to that placed on the medium variance likelihood (although the change is in the predicted direction). That said, our results are not simply more consistent with a rote process, since the weight placed on the high likelihood does not decrease with increasing exposure (no interaction between likelihood and block).

Table 2.2

Results from a 2 (prior) x 3 (likelihood) x 2 (block) Repeated Measures ANOVA for all likelihood variances in Experiment 1

	F	*df, df_{error}*	*p*	Effect size (η_p^2)
Likelihood	50.08	2, 50	<.001	.67
Prior	15.52	1, 25	.001	.38
Block	1.21	1, 25	.28	.05
Likelihood x prior	2.39	1.62, 40.41	.10	.09
Likelihood x block	.75	2, 50	.48	.03
Prior x block	4.35	1, 25	.05	.15

Our results point mostly away from a simple variance weighted Bayesian model being a good model of human behaviour in this particular task. The correct pattern of weights was present, but evidence of transfer was weak. Participants were also significantly sub-optimal, overweighting the likelihood whenever its variance was medium or high. Previous studies have also found that observers give more weight to the sensory cue than is optimal (e.g., Bejjanki et al., 2016); even so, the level of sub-optimality that we observe here is still drastically higher, compared to previous reports. However, Sato and Kording (2014) found better, near-optimal performance in those participants who were told that the sensory information can have one of two levels of variance, and that the variance will sometimes change, compared to those who were not provided with this information. We therefore reasoned that if observers are given additional information about the structure and statistics of the task (e.g., that the variances of the prior distributions are different), the weight they give to the sensory cue may move closer to optimal. If we find weights closer to optimal, we may be better able to detect whether transfer had taken place because the effect size of a change in the likelihood would be bigger. In fact, we wonder whether the size of this effect could be an important factor behind the lack of significant differences observed in Experiment 1, i.e., that the effect size of the change from medium to high likelihood was too small for our statistical analysis to reliably detect. In view of this, we set out to test whether additional instructions will lead to weighting of likelihood and prior information that is closer to optimal.

2.3 Experiment 2: Additional instructions about prior variance

Experiment 2 was identical to Experiment 1 except for a change in instructions. In this experiment, subjects were explicitly (albeit indirectly) informed of the different variances of the prior. We hypothesised that giving participants additional information about the model structure of the task will move weights closer to optimal and make any transfer effects more pronounced.

2.3.1 Methods

Twelve participants (8 female, mean age: 20.3, age range: 19-22 years) participated in Experiment 2. All participants had normal or corrected-to-normal visual acuity, no history of neurological or developmental disorders and had not taken part in Experiment 1. Each participant received either course credits or cash compensation for their time.

The experimental set had the same layout as the main experiment, with the following difference: in addition to the previously described instructions, subjects in this version of the task were told that "it is important to remember that one of the octopuses tends to stay in a particular area, whereas the other one moves quite a bit!" (i.e., they were indirectly informed that the variances of the priors were different). (see Appendix A for full instructions).

2.3.2 Results and Discussion for Experiment 2

Similarly to what we saw in Experiment 1, the mean and robust average of the dots explained the same amount of the variance in participants' responses ($R^2 = 0.991$ for both), followed by the median ($R^2 = 0.990$) and the mid-range ($R^2 = 0.989$). We, thus, proceed with the mean as the estimate from the likelihood.

Figure 2.3 shows that the pattern of results was qualitatively similar to those of Experiment 1 (see Figure 2.2). A 2 (prior) x 2 (likelihood) x 5 (block) repeated measures ANOVA (analysing only the low and medium likelihood trials) revealed that subjects placed less weight on the likelihood as its uncertainty increased (main effect of likelihood, $p < .001$) and as the prior uncertainty decreased (main effect of prior, $p = .002$). However, unlike in Experiment 1, there was no significant interaction of these factors ($p = .123$) (see Table 2.3).

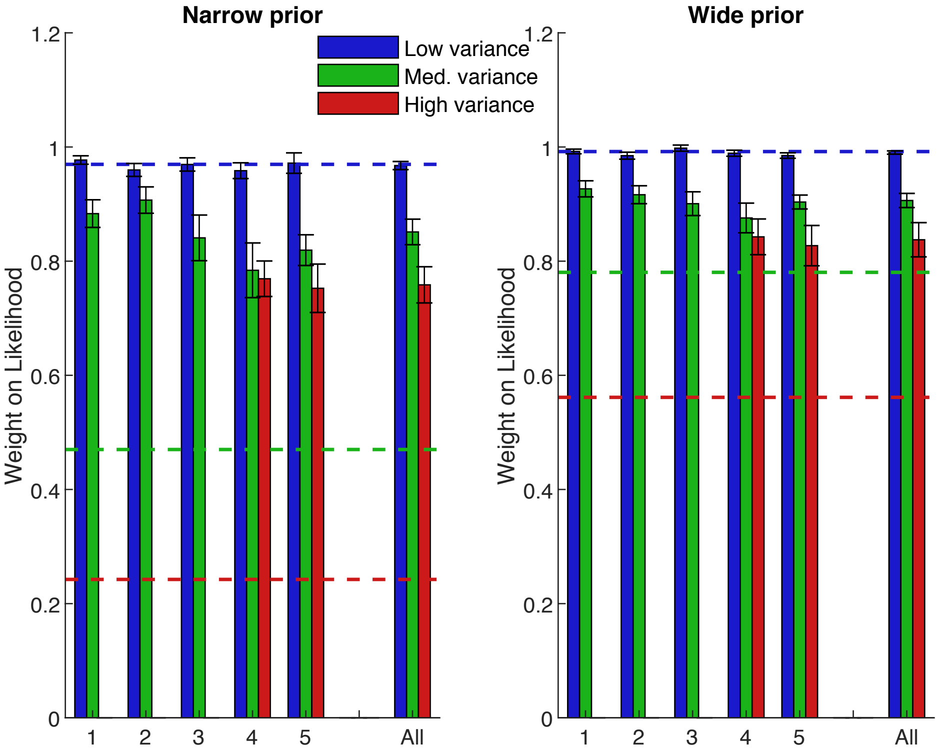

Figure 2.3. Mean weight placed on the likelihood information in each block of Experiment 2. Blue is the low-variance likelihood, green is the medium-variance likelihood, red is the high-variance likelihood. Dashed lines show optimal values. Error bars are +/- 1 SEM. The far right is the average over blocks.

We found a main effect of block ($p = .02$) and an interaction between block and likelihood ($p = .01$), with participants weighting the likelihood significantly less with increasing task exposure (regardless of prior) when its variance was medium ($F(2.21,25.34) = 3.81, p = .03, \eta_p^2 = .257$, with a Greenhouse-Geisser correction), but not when it was low ($F(4,44) = 0.70, p = .60, \eta_p^2 = .060$).

Table 2.3

Results from a 2 (prior) x 2 (likelihood) x 5 (block) Repeated Measures ANOVA for the likelihood variances present from the beginning (low and medium) in Experiment 2

	F	*df, df*$_{error}$	*p*	Effect size (η_p^2)
Likelihood	46.14	1, 11	<.001	.81
Prior	17.28	1, 11	.002	.61
Block	3.15	4, 44	.02	.22
Likelihood x prior	2.80	1, 11	.12	.20
Likelihood x block	3.73	4, 44	.01	.25
Prior x block	1.32	2.07, 22.77	.29	.11

As before, we analysed subjects' responses in the prior-only trials, finding a good quantitative agreement with the true prior mean (narrow prior: $t(11) = -0.002, p = .999;$ wide prior: $t(11) = 0.35, p = .734$). The median standard deviation (SD) of responses was also remarkably similar to the true prior SDs (narrow prior: 1.6% vs. 1.3% in screen units; wide prior: 2.5% for both).

Again, subjects' overall performance was suboptimal (as can be seen by comparing the height of the bars against the dashed lines – the optimal predictions – in Figure 3). Subjects' placed more weight on both the medium and high variance likelihoods than is optimal ($p < .001$ in both cases, for both priors). However, it is worth noting that the weights placed on the medium and high likelihoods are closer to optimal than they were in Experiment 1 (compare bar heights in Figure 2.2 and 2.3).

Finally, we tested for transfer to the newly-introduced high likelihood in blocks 4-5. We conducted a 2 (prior) x 3 (likelihood) x 2 (block) repeated measures ANOVA (analysing only blocks 4 and 5 with all likelihoods present). These results are shown in Table 2.4 and summarised here. There was a main effect of likelihood, with less weight placed on the likelihood as it became more uncertain ($p < .001$). Unlike in Experiment 1, post-hoc analysis showed that the weight placed on the high likelihood was significantly lower than the weight placed on the medium likelihood ($p = .034$). The weights placed on the likelihood in the low variance trial type were significantly lower than those in the medium and high variance trial types ($p < .001$ for both). Moreover, there was no main effect of block ($p < .64$), or an interaction effect of block and likelihood ($p < .15$), meaning that the weight placed on the newly-added likelihood information did not vary with increasing exposure.

Again, we find a significant difference between subjects' weights (in block 5) and optimal predictions when the likelihood variance was medium or high ($p < .001$ in both cases), but not when it was low, irrespective of prior variance (low likelihood, narrow prior: $t(11) = .120, p = .907$); low likelihood, wide prior: $t(11) = -1.29, p = .163$).

In line with our prediction of transfer, here we show that the observers put lower weight on the high variance likelihood than the medium variance likelihood they have experienced before. This is strengthened by the fact that participants' weights did not change significantly with increasing exposure across blocks 4-5.

Table 2.4

Results from a 2 (prior) x 3 (likelihood) x 2 (block) Repeated Measures ANOVA for all likelihood variances in Experiment 2

	F	*df, df_{error}*	*p*	Effect size (η_p^2)
Likelihood	37.98	2, 22	<.001	.78
Prior	18.36	1, 11	.001	.63
Block	.23	1, 11	.64	.02
Likelihood x prior	1.80	2, 22	.19	.14
Likelihood x block	2.09	2, 22	.15	.16
Prior x block	.17	1, 11	.69	.02

To check more directly for differences due to experimental instructions, we compared subjects' performance in the last 2 blocks across the two experiments. We ran a 2 (instructions) x 2 (prior) x 3 (likelihood) x 2 (block) mixed ANOVA. We found a main effect of instructions ($F(1,36) = 6.78, p = .013, \eta^2_{partial} = .159$) with subjects weighting the likelihood significantly less with explicit instructions (Experiment 2), i.e. closer to the optimal weightings. We also found an interaction between instructions and likelihood ($F(1,36) = 4.79, p = .011, \eta^2_{partial} = .118$), indicating that the main effect of instructions is due to a significant decrease in the weight placed on the high, relative to the medium likelihood in the explicit instructions (Experiment 2) task, but not the original task (Experiment 1).

These results show that adding extra instructions to the task that make the participant aware of a change in uncertainty between the two priors has an effect. The weights placed on the likelihood moved closer to optimal, and the transfer criterion was met, which suggests that, perhaps, observers are more likely to adopt a Bayes-optimal strategy when more explicit expectations about the correct model structure of the task are set. However, even with the additional instructions, the weight given to the sensory cue was still systematically higher than the "optimal" weight. Arguably, expecting people to perform optimally is rather unrealistic, as it presumes that the observer perfectly knows the environmental statistics. However, Bejjanki et al. (2016) found performance much closer to optimal than what we have seen in either of our previous experiments. The major difference between their experiment and ours' is the fact that Bejjanki et al. (2016) presented all likelihood variances from the start of the task. Therefore, unlike Experiments 1 and 2, which were designed in order to provide some evidence of transfer, Experiment 3 sought to test whether subjects' weights would move closer to optimal if we present all likelihood variances from the beginning, in a more direct replication of Bejjanki et al. (2016). The likelihood and prior variance parameters were identical to those used in Bejjanki et al. (2016), and we used a similar number of trials per prior and likelihood pairing (250 vs. 200 trials in Bejjanki et al., 2016).

2.4. Experiment 3: All likelihoods from the beginning

Experiment 3 was identical to Experiment 1 (lacking the extra instructions of Experiment 2) except that all likelihood variances were included from the beginning of the task. The participants experienced all six trial types in every block.

2.4.1 Methods

Twelve participants (10 female, mean age: 22.6, age range: 19-30 years) took part in Experiment 3. All participants had normal or corrected-to-normal visual acuity, no history of neurological or developmental disorders and had not taken part in Experiment 1 and 2. Each participant received either course credits or cash compensation for their time.

The stimuli and task were identical to those described for Experiment 1, except that all likelihood conditions (low, medium and high) were now present from the beginning (50 trials of each likelihood/ prior pairing interleaved in the same block).

2.4.2 Results and Discussion of Experiment 3

Again, the mean position of the dots explained the most amount of variance in participants' responses ($R^2 = 0.990$). The amount of variance explained decreased for the robust average ($R^2 = 0.989$), median ($R^2 = 0.988$) an the mid-range of the dots ($R^2 = 0.985$). We, thus, proceed with the mean as the estimate from the likelihood.

Figure 2.4 shows a similar pattern of results to Experiments 1 and 2. Again, a 2 (prior) x 3 (likelihood) x 5 (block) repeated measures ANOVA shows that the likelihood information was weighted less as it became more unreliable (main effect of likelihood, $p < .001$). Specifically, subjects placed significantly more weight on the low likelihood than on the medium ($p = .001$) or high likelihood ($p < .001$), and more weight on the medium likelihood than the high likelihood ($p = .005$). No other main effects or interactions were significant (see Table 2.5 for a summary of results).

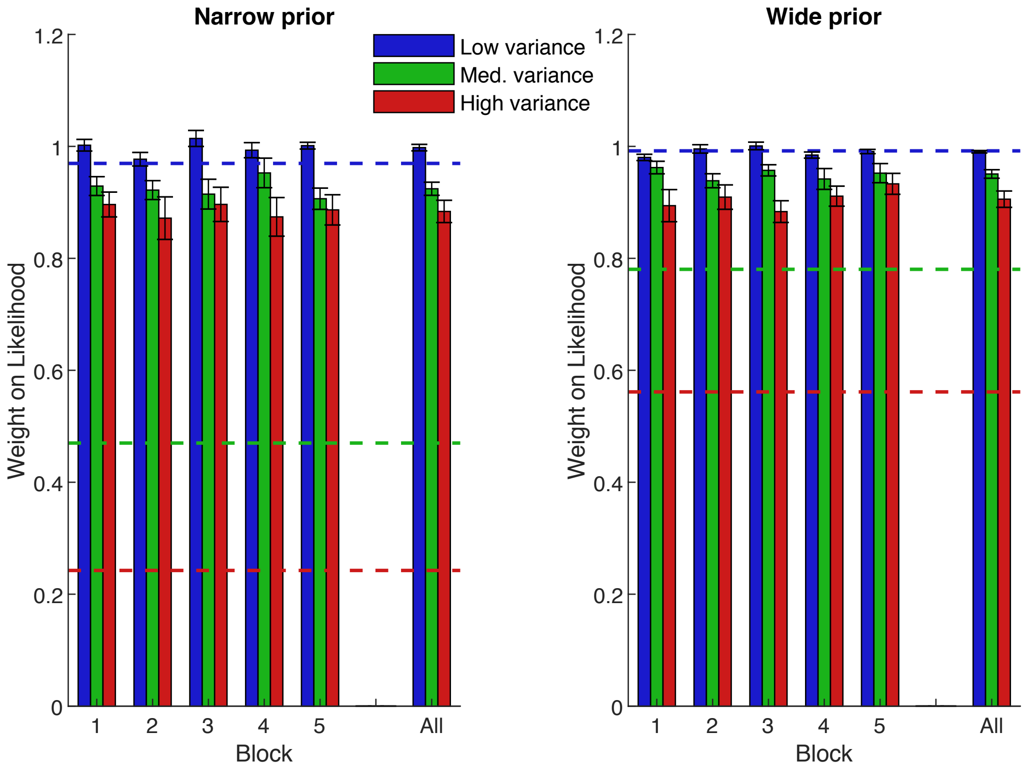

Figure 2.4. Mean weight placed on the likelihood information in each block of Experiment 3. Blue is the low-variance likelihood, green is the medium-variance likelihood, red is the high-variance likelihood. Dashed lines show optimal values. Error bars are +/- 1 SEM. The far right is the average over blocks.

Table 2.5

Results from a 2 (prior) x 3 (likelihood) x 5 (block) Repeated Measures ANOVA for all likelihood variances in Experiment 3

	F	*df, df*$_{error}$	*p*	Effect size (η_p^2)
Likelihood	29.90	2, 22	<.001	.73
Prior	2.74	1, 11	.13	.20
Block	.28	4, 44	.89	.03
Likelihood x prior	2.67	2, 22	.09	.19
Likelihood x block	.52	8, 88	.84	.05
Prior x block	.87	4, 44	.49	.07

For the prior-only trials, subjects' responses were, on average, statistically indistinguishable from the mean of the wide prior distribution ($t(11) = -1.14, p = .278$), but were significantly different from the mean of the narrow prior ($t(11) = -3.91, p = .002$) (although we note that the bias was small (95% CI: [0.24,0.87] percent of the screen width to the left). The median standard deviation (SD) of responses was 2.2% for the narrow prior condition and 2.6% for the wide prior condition; the SD of responses was, therefore, only close to the true variance of the wide prior (which was 2.5%). Together, these findings suggest that subjects had not learnt either the mean, or the variance of the narrow prior condition. This may explain the lack of difference in performance between the narrow and wide prior conditions in this task.

A comparison of subjects' weights on the likelihoods in block 5 against Bayesian predictions showed a significant difference for all likelihood and prior pairings ($p < .001$), with the exception of the wide prior/ low likelihood condition ($t(11) = -.362, p = .724$).

To sum up, although the correct pattern of weights was present, subjects were still substantially sub-optimal, even after experiencing all likelihood variances from the start of the task.

Accounting for Suboptimality

Even when we replicate Bejjanki et al. (2016) very closely with all likelihoods from the beginning, our participants are strikingly suboptimal. However, we note that in our initial calculations of optimal behaviour, we assumed that observers weight sensory and prior information according only to the variance of the dot distribution (i.e., external noise). However, many of the studies in the cue combination field that have found near-optimal performance used cues that only have internal noise, not external (Alais & Burr, 2006; Körding & Wolpert, 2006). It is, therefore, possible that our participants are sub-optimal because they fail to take account of external noise, only weighting the sensory and prior information by the internal variability (i.e., error intrinsic to them) of the sensory cue. Keeping this is mind, we considered the predicted weights of a model that only takes into account the internal variability σ_{l_i} in using the sensory cue.

We performed a separate control experiment to see how good participants were at finding the centroids of dot clouds in the absence of prior information (see Appendix B for more details). From this control data, we could calculate observers' internal noise as their responses were not subject to bias from the prior. For each participant, their internal variability σ_{l_i} was calculated by taking the standard deviation of their errors from the centroid of the dots (error = dot centroid - response). The predicted weight on the sensory cue was then calculated as $w = \sigma_p^2/(\sigma_{l_i}^2 + \sigma_p^2)$. This equation was the same as the full Bayesian model, the only difference being that the external SD of the likelihood (as defined by the experimenters) was substituted for the internal SD of the likelihood (measured in the control experiment). The variance of the prior was still included in the model.

We compared subjects' weights (block 5, Experiment 1) with those predicted when only weighting by internal noise and found that they were significantly different for all likelihood and prior pairings ($p < .01$). Indeed, Figure 2.5 shows that the internal noise model still predicts less weight on the sensory cue than we see in our data (compare bars and dotted lines). This could reflect participants downweighing the prior because it is, in fact, subject to additional internal noise, stemming from a need to remember and recall the correct prior from memory. Even so, empirical weights were closer to the internal noise predictions, compared to those, predicted by the optimal strategy (with experimentally-controlled cue variance, dashed lines).

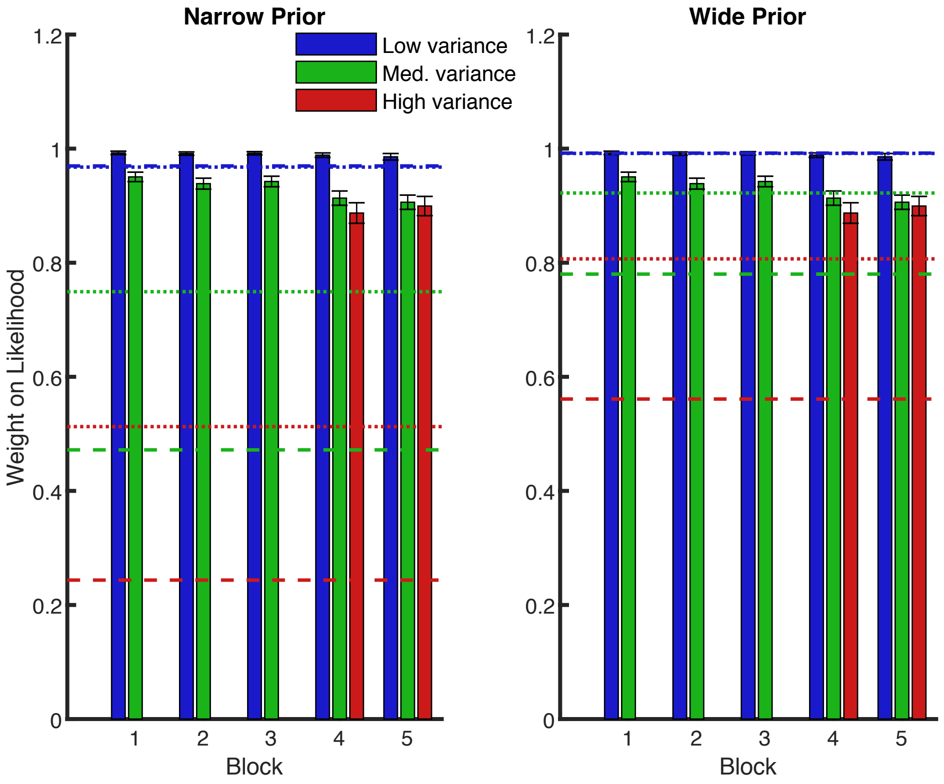

Figure 2.5. Mean weight placed on the likelihood information in each block of Experiment 1. Blue is the low-variance likelihood, green is the medium-variance likelihood, red is the high-variance likelihood. Dashed lines show optimal values. Dotted lines show predicted weights when only weighting by internal noise. Error bars are +/- 1 SEM.

In addition, we examined the predictions for an observer model that weights sensory information, according to overall variability in the sensory cue. When calculating the optimal predicted weights initially, we assumed that the optimal observer knew how reliable the dots (i.e., the likelihood) were, and could average them perfectly. Since participants will not be perfect at averaging dots, they will be more variable in using the sensory cue than the optimal observer. Therefore, the truly optimal thing to do is for participants to weight the sensory cue, according to their overall variability by taking into account both the variance of the dot distribution and the internal variability in estimating the average of the dots. Since the sensory cue is now less reliable (due to the added internal variability), we would expect participants to put less weight on it, and more weight on the prior.

We calculated the overall variability in using the sensory cue as:

$$\sigma_{l_o}^2 = \sigma_l^2/n + \sigma_{l_i}^2$$

where σ_l^2/n is the external noise in the sensory cue, and $\sigma_{l_i}^2$ is the individual internal variability.

As is expected, Figure 2.6 shows that the predicted weights in this case were lower than the optimal weights (compare dotted and dashed lines) as participants are worse than the optimal observer in averaging the dots. They placed less weight on the sensory cue and more weight on the prior than the optimal observer. We also compared these predicted weights to subjects' weights in the final block (5) in Experiment 1, and found that they were still significantly different from the empirical data when the variance of the likelihood was medium or high (irrespective of prior variance) and when the likelihood variance was low and the prior variance was narrow (all $p < .001$). No significant difference was observed when the likelihood variance was low, and the prior variance was wide ($p = .79$). This means that accounting for the added internal variability fails to explain our results as observers are placing more weight on the sensory cue than is optimal, not less.

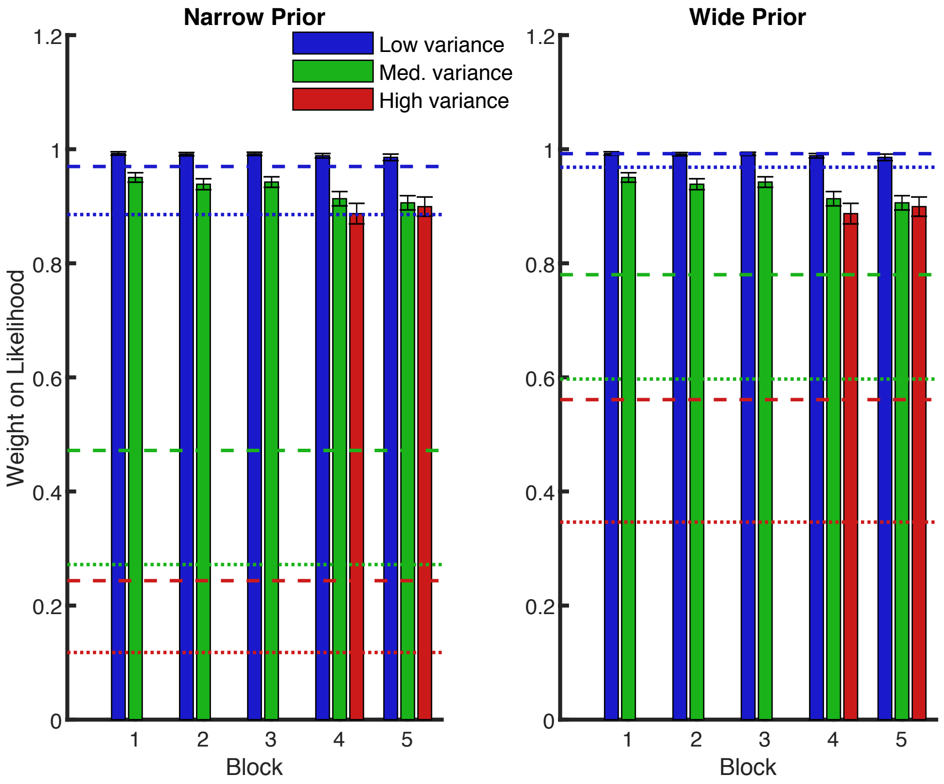

Figure 2.6. Mean weight placed on the likelihood information in each block of Experiment 1. Blue is the low-variance likelihood, green is the medium-variance likelihood, red is the high-variance likelihood. Dashed lines show optimal values. Dotted lines show predicted weights when overall variability in using the likelihood is taken into account. Error bars are +/- 1 SEM.

We compared the mean squared error (MSE) for each of the three models we tested: a) the original optimal model (using the experimentally imposed likelihood variance); b) the model with only the internal noise; and c) the model with the overall variability (including both the experimentally imposed likelihood variance and the internal noise). The internal noise model had the lowest MSE, which confirms that this model provides a better explanation for subjects' behaviour than other models (see Figure 2.7).

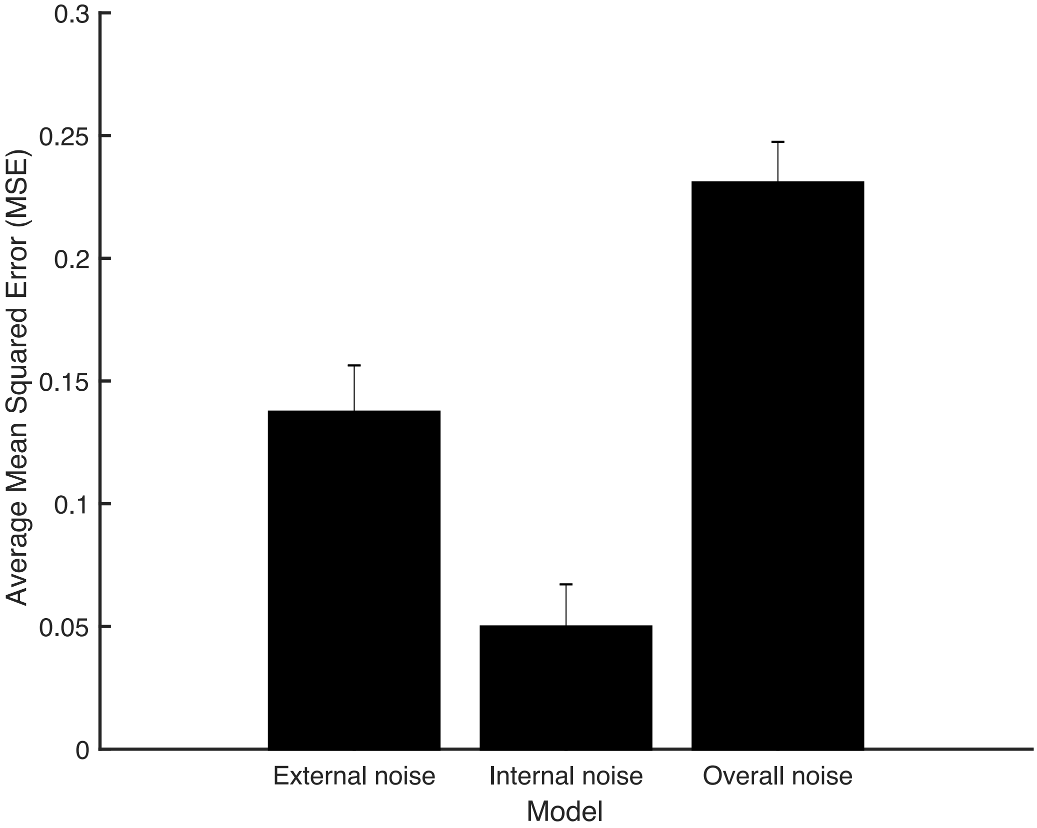

Figure 2.7. Average Mean Squared Error (MSE) for the external noise, internal noise and overall noise models. MSE calculations were based only on participants, for whom control data was available (N = 12; 6 had participated in Experiment 2 and 6 had participated in Experiment 3).

In summary, our data are best described by a model based on subjects' internally generated noise, as opposed to either a model with the experimentally imposed likelihood variance, or a model that accounts for both the experimentally imposed likelihood variance and the internal noise.

2.5 Discussion

We set out to test more strictly than before for Bayes-like combination of prior knowledge with sensory information in the context of a sensorimotor decision making task (Beierholm et al., 2009b; Bejjanki et al., 2016; Berniker et al., 2010; Tassinari et al., 2006; Vilares et al., 2012) by adding transfer criteria to the task (Maloney & Mamassian, 2009).

In Experiment 1, we did this by investigating whether observers are able to learn the variances of two prior distributions, and instantly integrate this knowledge with a new level of sensory uncertainty added mid-way through a task. We found that observers placed more weight on the sensory cue (likelihood) when its variance was low and the variance of the prior was high; behaviour that is in broad agreement with the Bayesian prediction. However, we found only partial evidence of transfer. The weight placed on the high variance likelihood was not significantly lower than that placed on the medium variance likelihood, which is at odds with the prediction of transfer. Importantly, even though qualitatively, our participants behaved like Bayesian observers, their performance fell markedly short of optimal.

In two further experiments we asked: (1) how behavior would be affected by additional instructions, which can clarify whether this suboptimality stems from using the incorrect model structure of the task; (2) whether experiencing the high likelihood variance condition from the start of the experiment would lead to closer-to-optimal weighting of the prior and likelihood information. In the first of these two further experiments, Experiment 2, we found that subjects' performance moved closer to optimal when they were (indirectly) instructed that the prior variances were different – possibly why we were able to detect significant evidence of transfer in the task. However, they were still significantly sub-optimal in multiple experimental conditions. Participants remained significantly sub-optimal in the final experiment (Experiment 3), when the need for transfer was removed (all trials types were present from the start of the task) and the experiment became a more direct replication of (Bejjanki et al., 2016).

Suboptimal weighting of prior and likelihood information

We show that observers take uncertainty into account, giving more weight to the sensory cue as its variance decreases, a result that is consistent across all three experiments. Equally, for a Bayes-like observer, we expect to find that the weight on the sensory cue is higher as the prior variance increases, but we found a main effect of prior in Experiments 1 and 2 only, and not in Experiment 3. Moreover, while our manipulation to the instructions in Experiment 2 moved the weights placed on the likelihood closer to optimal, they were still significantly different to the optimal predictions.

To examine to what extent additional sensory variability in estimating the centres of dot-clouds could have affected predictions and performance, we ran a separate, control experiment (see Appendix B). This shows that observers are less efficient in their use of the likelihood information than an ideal observer: the variability of their responses is significantly larger than the true variability of the sensory cue in both the low and medium variance likelihood conditions. However, this fails to account for suboptimal performance: ideal weights for the likelihood that are computed using the measured likelihood variabilities in the control task are still significantly lower than those in the empirical data.

Suboptimal weighting of the prior and likelihood information may also be caused by incomplete or incorrect learning of the prior information. However, the prior-only trials suggest that the observers learn the means of the priors and distinguish between their variances at least in Experiments 1 and 2, if not Experiment 3 (under the assumption that standard deviations of subjects' responses are related to the learnt prior variances). Suboptimal weighting of the prior could also be due to the use of an incorrect Bayesian generative model (causal structure) by subjects, e.g. if they believe that the prior will change over trials then they should apply a smaller weight to the prior (could be conceptualized as a meta-prior or hyperprior in Bayesian terms, Gelman, Carlin, Stern, & Rubin, 2013). The fact that we found an effect of instructions imply that the causal structure assumed by subjects can indeed greatly influence behavior (Shams & Beierholm, 2010).

Other research groups have performed similar experiments but using only a single prior distribution (Acerbi et al., 2014a; Berniker et al., 2010; Chambers et al., 2018; Tassinari et al., 2006; Vilares et al., 2012). These studies also find deviations from the optimal predictions; however, the deviations can be accounted for by adding extra sources of inefficiency to the model that are due to motor errors, centroid calculation errors, and aim point (in reaching tasks) calculation errors (Tassinari et al., 2006). Moreover, when trials are blocked by prior condition, it has been shown that learning after a switch in the prior variance is slower when the prior variance decreases than when it increases, suggesting participants may perform optimally after further exposure to the task (Berniker et al., 2010).

We considered elements of the experimental design that could have resulted in suboptimal behavior in a task similar to others in the literature where performance was closer to optimal (e.g., Bejjanki et al., 2016). Firstly, describing the dots as the tentacles of an octopus may have caused participants to assume that another method of gaining an estimate from the dot colour, other than taking the mean horizontal position, was more appropriate in this task (de Gardelle & Summerfield, 2011; Van Den Berg & Ma, 2012). However, our analysis shows that participant responses are not better predicted by the median, mid-range, or robust average, than they are by the mean. Secondly, our correction of the dot positions so that their SD on each trial was equal to the true likelihood SD may have influenced participant's inferred reliability for the likelihood. However, an observer who computes the reliability for the likelihood trial by trial, by taking the dot cloud SD, would infer that the likelihood was less reliable ($\frac{1}{\sigma_l^2}$) as a cue to true location than the centroid of the dots would be for an observer who could perfectly calculate the mean of the dots ($\frac{1}{\sigma_{l/n}^2}$). This would lead to an observer placing less weight on the likelihood than the ideal observer. Participants in our experiment place more weight on the likelihood than the ideal observer, so this is not the source of suboptimality in our experiment.

Finally, whilst the true likelihood and prior reliabilities used in our task were matched to those in Bejjanki et al. (2016) observers may have perceived the cue (dots) as more reliable than it actually was, which in turn would result in more weight placed on the cue than in previous studies (Bejjanki et al., 2016; Vilares & Kording, 2011).

It is possible that subjects did not experience enough trials of each prior and likelihood uncertainty to reach optimal performance, and indeed, we find evidence of decreasing weights on the likelihood with increasing task exposure in both Experiments 1 and 2 (main effect of block, although not the rise from block 4 to 5 in Experiment 2). Crucially, however, our participants experienced more trials per prior than in Bejjanki et al. (2016) (750, compared to 600) where weights were closer to optimal, ruling out the possibility that observers did not experience enough trials to learn the complex features of the distributions.

The result that our participants' performance was so different -in terms of level of sub-optimality- compared to Bejjanki et al. (2016) might be explained by a difference in instructions. Specifically, their instructions have a social element that ours do not, i.e., in their task, participants were instructed to interpret the likelihood dots as “locations that other people have previously guessed the bucket is located at” (vs. tentacles of the octopus in ours). This means that in Bejjanki et al. (2016) participants would have to take into account how accurate they think other people’s guesses are. If we assume that people give lower weight to information that is allegedly based on other people’s guesses, this might explain why observers in Bejjanki et al. (2016) generally weighted the likelihood less than in our experiments (Martino et al., 2017). Another aspect about the instructions that is worth mentioning here is that, perhaps, our participants are more likely to assume that the body of an octopus is in the center of its tentacles, compared to previous guesses of other participants (Bejjanki et al., 2016) or splashes from a coin (Tassinari et al., 2006; Vilares et al., 2012). However, had this been the case, we would have expected participants’ responses to be better predicted by another estimate, such as the robust average, than the mean of the dots, and we found no evidence of this in the data.

Another explanation is that observers were being “optimally lazy”; that is, they deviated from optimal performance in a way that had minimal effects on their expected reward (Acerbi, Vijayakumar, & Wolpert, 2017). In this case, we would expect obtained reward to match well with the predictions of the optimal Bayesian model; instead, the predicted reward resulting from optimally combing sensory and prior information was higher than that obtained by our observers - particularly when the variance of the prior was narrow (Appendix C). Therefore, we have no reason to believe that the suboptimal behaviour we observed in our task was due to our participants being “optimally lazy” (Acerbi et al., 2017).

Nevertheless, we could show that the suboptimal behaviour in our task can be best explained by assuming that participants were weighting sensory information (relative to prior information) only according to the internal variability in using the cue, ignoring external noise. It is, thus, interesting to consider it as one potential explanation, on the computational level, for the deviations from optimal consistently reported in similar studies on combination of sensory and prior information (Bejjanki et al., 2016; Berniker et al., 2010; Sato & Kording, 2014; Tassinari et al., 2006; Vilares et al., 2012). However, we note that attending to internal noise may be easily mistaken for underweighting the total noise; future work could investigate the extent to which suboptimal behaviour is specifically linked to the use of internal variability, and not simply a general under-estimation of the total noise in the stimuli.

Note that the observer model, based on the internally generated noise, can still be considered “subjectively” optimal (fully Bayesian), in the sense that observers take into account and act according to their internal noise variability (Acerbi et al., 2014a). This strategy looks sensible but is arguably not Bayes-optimal as an ideal observer has to take into account external sources of noise in addition to his or her own sensory uncertainty (Kersten et al., 2004; Knill & Richards, 1996).

Evidence for Bayesian transfer

We found only partial evidence for transfer in Experiment 1, as there was no significant change in the weight placed on the likelihood between the medium and high likelihood conditions. In fact, subjects seemed to treat the high variance likelihood the same as the medium variance likelihood (that they had experience with), suggesting that observers did not adopt a statistically optimal Bayesian strategy. Nonetheless, performance did not improve with more trials, suggesting that subjects were not implementing a look-up table decision rule, either (Maloney & Mamassian, 2009). However, we note that in our data, observers placed much more weight on the medium and high likelihoods than is optimal. This means that the effect size of a change in likelihood is much less than was expected; thus, the observed lack of significant differences might simply be due to lack of statistical power in our analysis to detect such small effect sizes.

Why do we see more convincing evidence of transfer in the instructions task? Bayes-like computations demand considerable computational resources (e.g., working memory load, attentional focus); it is, therefore, reasonable to expect that if a task is sufficiently complex, and there is a lot to learn, subjects will start behaving sub-optimally. The impact of additional instructions in Experiment 2 may be to free up cognitive resources by providing subjects with (indirect) information about the variances of the two prior distributions at the start of the task (Ma, 2012; Ma & Huang, 2009).

Our findings do not allow us to clearly distinguish between the reinforcement-learning and Bayesian interpretations. We found that when we introduced a new level of (known) uncertainty to the likelihood, observers immediately changed how they used this new information in a way that is largely consistent with optimal predictions; this effect was significant in the second experiment, but not in the first. Thus, we note that this effect is not particularly robust as it depends on the experimental procedure used to measure it. Indeed, our findings demonstrate that whether this effect is observed in the first place is greatly affected by small changes in experimental layout (e.g., instructions, number of trials). The fact that we observe no learning during Experiment 3 (no main effect of block), coupled with the observation that the weight on the likelihood in the 4th block of Experiment 3 were remarkably similar to those in the 4th block of Experiments 1 and 2 makes a weak suggestion of a Bayesian interpretation. However, a stronger test of transfer would be if participants had received no feedback for the new level of uncertainty. We provided trial-by-trial feedback (true target position) to ensure that participants were able to learn and recall the correct prior distributions. Therefore, we cannot rule out the possibility that our participants used the feedback to directly learn a mapping between the high variance likelihood and each prior, instead of the distribution of locations.

Sato and Kording (2014) showed that subjects behave in a Bayes-optimal fashion in a sensorimotor estimation task, where they transferred their knowledge from the ‘learning phase’ to the prior in the testing phase in the absence of trial-to-trial feedback, suggesting that people did not learn a simple likelihood-prior mapping. This means that the features of our experiments set an approximate upper bound on learning; in other words, we can generally expect subjects’ performance to be less accurate when performance feedback is not provided.

Nevertheless, in order to meaningfully test whether observers can transfer probabilistic information across different conditions, an experiment where trial-by-trial feedback is limited, or excluded altogether, is needed. Hudson, Maloney, and Landy (2008) argued that providing only blocked performance feedback, for example, would prevent participants from using a “hill-climbing” strategy in the high variance likelihood condition (i.e., updating their estimates, based on the feedback from trial to trial). Alternatively, Acerbi, Vijayakumar, and Wolpert (2014) found that partial feedback (where participants are told whether they “hit” or “missed” the target, but the actual target position is not displayed) is sufficient to maintain participant engagement; however, no meaningful information can be extracted from the feedback, preventing participants from using it to better their performance. Future work could investigate how the removal of full performance feedback would affect behaviour in more complex scenarios.

What are observers if not Bayesian?

Some studies suggest that BDT is generally a good descriptive model of people’s perceptual and motor performance, but quantitative comparison shows divergence from Bayes-optimal behaviour (Bejjanki et al., 2016; Zhou et al., 2018), not unlike what we report in this study. These deviations from optimality may have arisen because rather than performing the complex computations that a typical Bayesian observer would do, observers draw on simpler non-Bayesian, perhaps even non-probabilistic, heuristics (Gigerenzer & Gaissmaier, 2011; Zhou et al., 2018). Laquitaine and Gardner (2018) developed a model that switched between the prior and sensory information, instead of combining the two, which was found to explain the data better than standard Bayesian models. The authors concluded that people can approximate an optimal Bayesian observer by using a switching heuristic that forgoes multiplying prior and sensory likelihood. In another study, Norton, Acerbi, Ma, and Landy (2018) compared subjects’ behaviour to the ‘optimal’ strategy, and well as several other heuristic models. The model fit showed that participants consistently computed the probability of a stimulus as belonging to one of two categories as a weighted average of the previous category types, giving more weight to those seen more recently; subjects’ responses also showed a bias towards seeing each prior category equally often (i.e., with equal probability). We note that a Reinforcement-Learning (RL) model was also tested, where participants could simply update the decision criterion after making an error with no assumptions about probability; no participant was best fit by the RL model. This suggests that observers are, in fact, probabilistic, i.e., take into account probabilities, though not necessarily in the optimal way; instead, they seem to resort to heuristic strategies. However, future work should explore which, if any, of these models can capture the behaviour on this type of complex localisation task.

2.6 Moving from Limits on Using Information Efficiently in Adulthood to Childhood

The question of whether humans making decisions under uncertainty approach optimal decision makers has led to much debate (Rahnev & Denison, 2018). One way to resolve it is to learn more about the exact circumstances in which observers do and do not behave like ideal decision makers. Here we show that the capability of adults to rapidly and flexibly perform Bayesian inference is limited in situations of greater complexity (see also Rahnev & Denison, 2018). Evidence for this can be seen from the fact that 1) performance was very far from optimal values, and 2) there was only weak evidence of generalisation to an untrained cue reliability (Experiment 1, Chapter 2). We have therefore concluded that increasing cognitive complexity is a key limiting factor for the implementation of Bayesian inference in adults.

What about children? We know that although children can extract information from environmental statistics within the first few months of life (Fiser & Aslin, 2002; Kirkham et al., 2002; Saffran et al., 1996), they do not begin to use them efficiently until 8-10-years of age (Bejjanki et al., 2019; Chambers et al., 2018). However, the factors that contribute to whether or not children use information efficiently (i.e., in a Bayesian, statistically optimal way) have received relatively little attention. Across three experiments in Chapter 3, we probe how different factors contribute to the less-efficient adaptation of young children's behaviour to novel environmental statistics.

Chapter 3

Development of efficient adaptation to novel task statistics

3.1. Introduction

A fundamental problem for human perception is processing the often uncertain and ambiguous information available to our senses. Luckily, our environment is highly structured in both space and time. For instance, the world tends to be stable over short time scales, and stimuli that have occurred recently are likely to reoccur (de Lange et al., 2018). Consequently, we can leverage information from the recent past to interpret incoming sensory signals (Beierholm, Rohe et al., 2020; Fritsche et al., 2017; Summerfield & De Lange, 2014).

Human adults can integrate their recent experience of the underlying structure of a new environment (prior experience, or *priors*), with current sensory inputs in a statistically (near-)optimal fashion, that is, weighting these information sources by their relative reliabilities (Bejjanki et al., 2016; Berniker et al., 2010; Vilares et al., 2012; Wolpert et al., 2011). However, little is known about development of abilities to use novel statistical regularities in this way. We know that infants already show a remarkable sensitivity to statistical regularities from an early age (Fiser & Aslin, 2002; Kirkham et al., 2002; Saffran et al., 1996) – for example, rapidly learning novel statistical regularities in spoken speech at 8 months (Saffran et al., 1996) and in visual patterns at 9 months (Fiser & Aslin, 2002). Being sensitive to novel statistics is crucial, but how effectively are these used to improve perception and decision-making? Studies comparing children's performance with adults and ideal observer models find that children apparently do not make adult-like use of new statistical regularities until age 9 years or later. A crucial test is whether participants shift their reliance (weighting) for prior vs. current sensory information as these change in their relative reliabilities. For example, when the prior becomes less reliable, the ideal observer would give more weight to the current sensory input. To test whether children weight a prior by its reliability, Chambers et al. (2018) varied the reliability of a visual cue and a prior provided to adults and 6- to 11-year-old children localising visual targets. The ‘prior’ here defines the distribution from which the target locations are sampled, and could be learned through experience with past target locations. Measuring weights given to each information source, Chambers et al. (2018) found that reliability-weighting increased with age: older children (> 9 years) adapted to changes in sensory uncertainty, as well as the uncertainty in the prior. By contrast, younger children were relatively insensitive to changes to the prior.

It therefore seems that although humans already detect and learn novel statistical regularities in infancy, it may not be until late childhood that they use these efficiently during perceptual tasks. What factors might underlie the long development of this critical perceptual skill? Another aspect of efficient information integration – reliability-weighted cue combination - also does not develop until the age of around 9 years (Dekker et al., 2015; Gori et al., 2008; Nardini et al., 2008, 2010), which suggests that there could be in part a central issue with representing and / or computing with probabilities. However, where learning and use of different statistical distributions is concerned, there may be additional, even more basic, task demands related to working memory and other executive functions, which develop markedly and over long time scales throughout childhood (Best & Miller, 2010; Diamond, 2013; Gur et al., 2012). One recent study suggests that, indeed, abilities to overcome such central cognitive demands may play a crucial role in development of abilities to make efficient use of novel task statistics (Bejjanki et al., 2019). However, research in this area is so limited at present that even the central issue of how children’s adaptation to task statistics may differ from adults’ is not yet clear.

In the present studies, we adopt a detailed approach to this problem. Building on previous studies which averaged performance over the experiment as a whole (Bejjanki et al., 2019; Chambers et al., 2018), we trace for the first time the time course of how children and adults adapt to novel, changing task statistics, as they are experiencing them. Considering the time-course of adaptation to task statistics lets us compare the learning rates and the end-points reached as a function of age, and to measure how groups adapt to changes in the prior distribution and other experimental manipulations.

In Experiment 1, we studied the time course of how children and adults adapt to underlying task statistics in a target-localisation task with a noisy sensory cue and a prior distribution that can be learned over the course of the experiment. In two follow-up experiments we probe two potential sources of slower tuning of children's behaviour to task statistics: i) limited working memory for representing the prior (previous locations) and ii) limited executive function abilities for learning and implementing multiple weighting rules simultaneously.

3.2 Experiment 1

We tested 6-8 year olds, 9-11-year-olds and adults on a spatial localisation task that reveals the time course of adaptation to task regularities (Berniker et al., 2010). Subjects were asked to estimate the hiding place of a target object by integrating information from a noisy visual cue and, potentially, prior knowledge (i.e., the distribution of previous target locations). The reliability of the prior was reduced halfway through the experiment in order to test whether this change in the prior would lead to re-weighting of the prior information, as in Berniker et al. (2010). We compared the relative weighting of prior knowledge and sensory input in adults and children aged 6 to 11 years. The age range was in accordance with previous studies looking at how children combine prior expectations and current sensory information (Bejjanki et al., 2019; Chambers et al., 2018), or multiple sensory cues (Gori et al., 2008; Nardini et al., 2010; Negen et al., 2019).

An ideal observer would assign more weight to a prior when it is more reliable (relative to a sensory cue), and less when it is less reliable. We would also expect observers to learn to use a prior with increasing experience. We asked whether adults assigned lower weights to less reliable priors overall (**H1**) and whether they changed their weight for each prior with increasing experience (**H2**). These initial tests serve as a baseline and a check that we replicated the expected results in adults. We next asked whether there were age differences in overall weights assigned to priors (**H3**), and in the degree of reweighting with increasing experience (**H4**). Finally, we asked whether either child or adult groups' weights for each prior were significantly different from optimal by the end of their experience with it (**H5**).

3.2.1 Methods

3.2.1.1 Overview. We adapted Berniker et al.'s (2010) task into a child-friendly, ‘octopus-catching’ game (Figure 3.1a). On each trial of the task, the true location of an octopus was sampled from a Gaussian distribution centred in the middle of the screen and with a narrow (first half of trials) or wide (second half of trials) variance. To accurately predict the location of the octopus, subjects could combine their prior expectations of where the octopus was likely to appear (that they could learn via trial-to-trial feedback) with a noisy sensory cue that changed from trial to trial. The cue was a single dot sampled from a Gaussian distribution centred on the true location with a fixed variance. To respond, participants moved a green vertical bar (i.e., the fishing net) horizontally to try to point as close as they could to where the octopus was hiding.

3.2.1.2 Participants. We tested adults (mean age = 21.82 years, age range 18-33 years of age; *n* = 19) and two age groups of children (6-8-year-olds: mean age = 7.71 years, *SD* = 0.89 years; *n* = 15; 9- 11-year-olds: mean age = 10.21 years, *SD* = 0.96 years; *n* = 17). All adult participants across Experiments 1-3 were recruited from the undergraduate psychology pool at Durham University and received course credit in return for study participation. Similarly, child participants in all three experiments were recruited through a database of local families who had agreed to be contacted for research studies and received a small toy as a thank you gift. In addition, stickers were used to incentivise actual performance in both adult and child groups. All participants had normal or corrected-to-normal vision.

3.2.1.3 Ethics. The study was approved by the University Ethics Committee, in accordance with GDPR regulations, and carried out according to the principles laid down in the 1964 Declaration of Helsinki. All participants (or their caregivers, as appropriate) gave informed consent in writing prior to taking part in the study.

3.2.1.4 Stimuli and Apparatus. Stimuli were displayed on a 15-inch Dell laptop (1920 x 1080 pixels), viewed at a distance of approximately 60 cm, using the

Psychophysics Toolbox for MATLAB (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). The background was set to blue, like the sea surface.

We define all stimuli in terms of screen units, where the left edge of the screen is mapped to 0 and the right edge is mapped to 1. In these units, the position of the octopus was sampled from a Gaussian distribution, centered on the middle of the screen ($\mu_p = 0.5$) with either a narrow (standard deviation (*SD*) of $\sigma_{pn} = 0.05$) or wide (SD of $\sigma_{pw} = 0.2$) variance; see Figure 3.1c. These parameters were identical to those used in Berniker et al. (2010). All participants experienced the narrow prior distribution first, followed by the wide prior distribution. This order was chosen because participants in Berniker et al. (2010) displayed a faster rate of learning when transitioning from narrow to wide. We therefore chose to focus solely on the narrow-to-wide change in variance as any developmental changes were more likely to be picked up in this order of presentation.

At the beginning of each trial, a single dot (diameter 0.05) appeared on the screen. The horizontal position of the dot was drawn from a Gaussian distribution centred on the true octopus' location with SD $\sigma_l = 0.1$). The likelihood SD remained fixed for the whole duration of the experiment. To respond, participants set the position of a vertical green bar (width 0.05, height 3% of screen height) on the horizontal axis. Feedback was given in the form of a red dot (diameter 0.05) that represented the octopus' true hiding place.

To allow us to determine whether adults in our study showed similar pattern of cue weights and learning rates to those in Experiment 2 in Berniker et al. (2010), the experiment included the same number of trials as those presented by Berniker et al. (2010). The experiment consistent of 500 trials, 250 trials for each prior variance condition: narrow, wide.

The experiment took, on average, 50 minutes to complete.

3.2.1.5 Procedure. Participants were told that there were octopuses hiding under the sea surface and their job was to estimate their location (Figure 3.1a). On each trial they saw a noisy visual cue, a 'bubble', after which they used the mouse to move a vertical green bar (i.e., a "fishing net") horizontally to try to click as close as they could to where the octopus was hiding. The visual cue remained on the screen until participants responded. Feedback, indicating the true location, was given on each trial along a score. Feedback is crucial in order for participants to learn the reliability of the prior and to detect when its reliability changes in the second half of the experiment. The maximum score was 20 points for a perfect “hit” and decreased as a participant’s response became further from the target. For a perfect “hit”, the net and the target needed to overlap by at least 50% of the red dot’s width (Figure 3.1b). When the participant picked the right location, we also incorporated an animation of a cartoon octopus smiling, with “+20” shown above it. If the target was within 70 pixels of where the participant responded, the participant received 5 points (“+5” was shown on the screen). If the target was further than 70 pixels away, the participant received no points, and a sad cartoon octopus was shown. Additionally, participants “levelled-up” every 100 points, after which they were shown a congratulatory screen and offered a sticker for completing the level. To ensure that both children and adults were interested in the stickers, each participant was asked to choose their favourite kind of sticker among three options. Participants were allowed to keep any stickers that they earned during the experiment. Progress towards the next level was always shown at the top of the screen, via a progress bar. The cumulative score was displayed at the end of each trial.

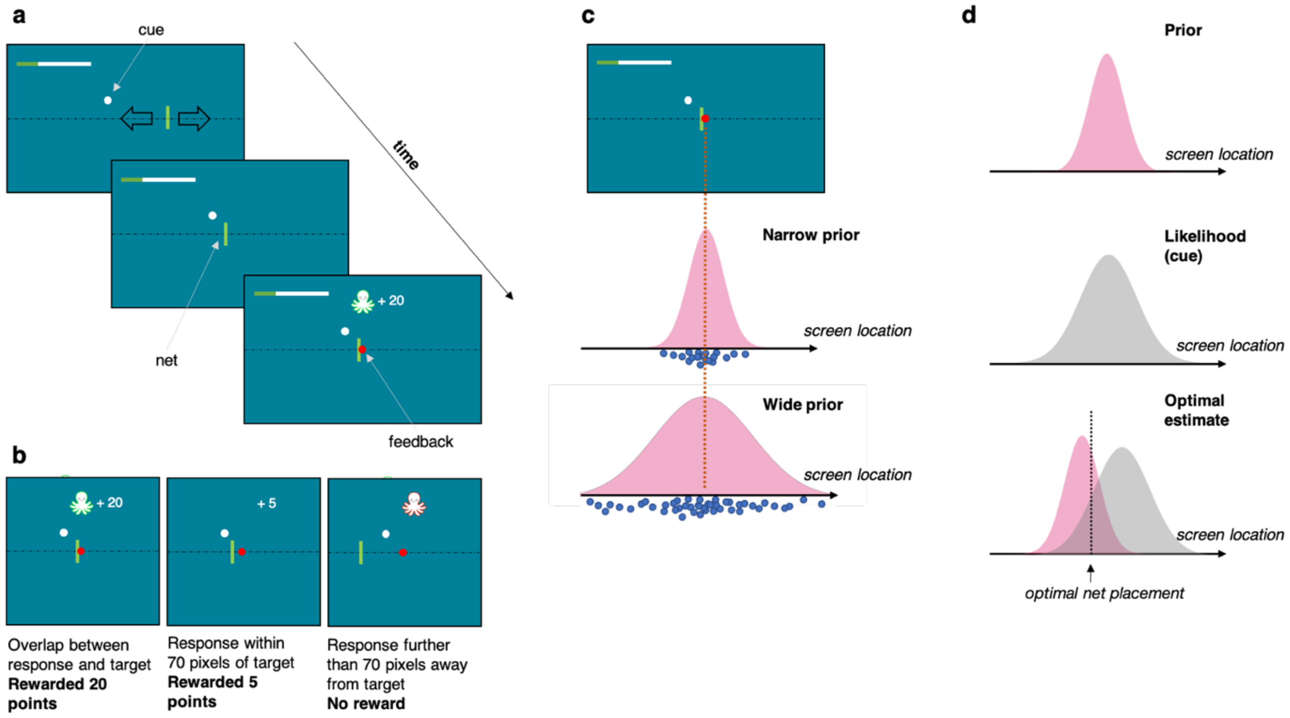

Figure 3.1. (**a**) An example of an experimental trial sequence. Each trial began with the presentation of a noisy visual cue to the target's current location, a single white dot described to participants as a "bubble" the octopus has blown on that trial. A net (vertical green bar) was presented at a random place on the screen at the same time. Participants estimated the hiding place of the octopus by adjusting the horizontal position of the net and pressing the mouse button to confirm their choice. Afterwards, feedback was given. Feedback consisted in a visual display of the true position of the target and a score that was maximal (20 points) for a perfect 'hit' and decreased away from the target (to five points for close but non-overlapping responses and no points for responses too far away from the target). (**b)** Reward structure: overlap between response and target location, resulting in 20 points (left panel); response within 70 pixels of target location, resulting in 5 points (middle panel); response further away from 70 pixels from target location, resulting in no reward (right panel). (**c)** SD of the prior. The true octopus' location (red dot) was drawn at random from a Gaussian distribution, centered on the middle of the screen and whose SD was 'narrow' in the first half of the experiment and 'wide' in the second half of the experiment. (**d**) Ideal observer prediction. According to BDT, a Bayesian ideal observer combines the prior distribution (the pink distribution) with the likelihood function to obtain the 'optimal' target location (black dotted line).

3.2.1.6 Analysis.

Calculation of weights

To allow us to test our hypotheses about weights given to the prior vs. sensory cue, an initial analysis was performed to calculate the weight assigned to the cue vs. the prior on each trial. To track how prior weighting may change over with time, for each participant, each consecutive twenty-five trials were binned together. This binned data was then used to perform linear regression of the estimated location (response) as a function of cue location. All regression analyses were done using a least squares procedure (the polyfit function in MATLAB). The slope of the regression fit is a measure of the weighting subjects placed on the cue, relative to the prior's mean, when estimating the target octopus' location. Slopes close to zero represent high reliance on the average prior target location, while slopes close to one represent high reliance on the sensory cue (i.e., low reliance on the prior). Hereinafter, we will refer to the fitted slope values as the weight placed on the sensory cue.

Flexibly integrating prior knowledge and the sensory cue

To test the hypotheses that adults assign lower weights to less reliable priors overall (**H1**) and change their weight for each prior with increasing experience (**H2**), we conducted a repeated-measures ANOVA with the weight on the cue (adults only) as the dependent variable and prior variance and bin as the within-participants factors. To test hypotheses about age differences in overall weights assigned to priors (**H3**), and in the degree of reweighting with increasing experience (**H4**), we conducted a mixed ANOVA on cue weights, with prior variance and bin as within-participants factors and age group as a between-participants factor.

Ideal observer predictions

To test whether any group's weights for each prior were significantly different from optimal by the end of their experience with it (**H5**; Figure 3.1d), we compared participant weights in the final bin of each prior to the weight predicted by the ideal, reliability-weighted averaging model for each age group separately using one-sample t-tests. The ideal (Bayesian) observer prediction is given by:

$$\omega_C = \frac{\frac{1}{\sigma_l^2}}{\frac{1}{\sigma_l^2} + \frac{1}{\sigma_p^2}} \qquad (1)$$

where ω_C is the ideal weighting of the likelihood, and σ_l^2 and σ_p^2 are the variances of the likelihood and the prior, respectively. In the current study, the variance of the likelihood was σ_l = 0.1 and the variance of the prior was σ_p = 0.05 (narrow variance) or 0.2 (wide variance). Substituting the reliabilities of the prior and likelihood into eqn. 1 shows that when prior variance was narrow, the ideal weight to place on the cue was 0.2.

$$\omega_C = \frac{\frac{1}{0.1^2}}{\frac{1}{0.1^2} + \frac{1}{0.05^2}} = 0.2 \qquad (2)$$

The ideal weight to give to the cue changed from 0.2 to 0.8 when the variance of the prior became wider.

3.2.2 Results

Figure 3.2 shows the weight assigned to the prior during the experiment (solid lines; see Calculation of weights). Participants were analysed in three age groups (6-8 years: upper panel, 9-11 years: middle panel, adults: bottom panel), consistent with previous studies that have reported group comparisons (Chambers et al., 2018). Also shown in Figure 3.2 is the weight that would be predicted for an ideal Bayesian observer (dashed lines). In summary, as described below, all age groups placed increasingly more weight on prior information when the prior was narrow (compare blue and yellow solid lines in all panels, Figure 3.2). However, the age groups differed in the extent to which they weighted the prior information: particularly when the prior was narrow, younger participants placed, on average, less weight on the prior than older children and adults as they performed more trials (see solid blue lines across all panels, Figure 3.2). Figure 3.2 also shows that 6-8-year-olds and 9-11-year-olds have shallower slopes than adults, suggesting that they adapted slower as we elaborate further below.

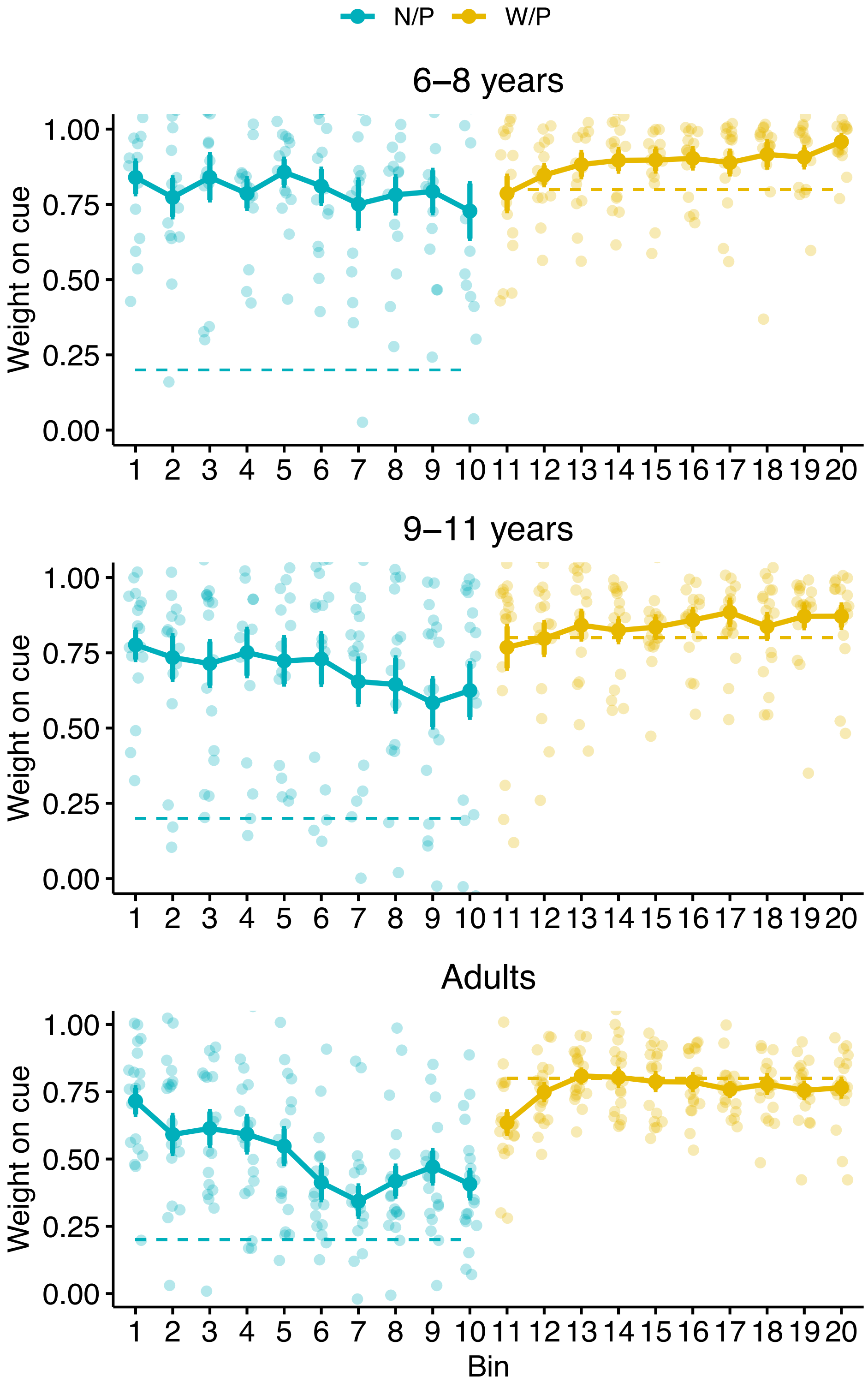

Figure 3.2. Mean weight placed on the cue during Experiment 1, separated by bin and prior variance (blue: narrow variance, yellow: wide variance) for each age group (6-8-year-olds: upper panel, 9-11-year-olds: middle panel, adults: lower panel). There are 25 trials in each bin. Lower values represent a greater weight on the prior.

In all panels, each circle is a participant, and error bars are standard errors of the mean (s.e.m). The dashed lines indicate optimal predictions (all panels).

We turn to the analyses addressing our hypotheses, beginning with those about the adult participants (Figure 3.2, bottom panel), for whom there is an expected pattern of performance based on the earlier Berniker et al. (2010) study using a similar method.

Adults assigned less weight to the less reliable prior overall.

To test whether adults would assign lower weights to the less reliable prior overall (**H1**), we entered adults' cue weights shown in Figure 3.2 (bottom panel) into 2 (prior variance: narrow vs. wide) x 10 (bin) repeated-measures ANOVA. Consistent with **H1**, we found a main effect of prior, $F(1{,}18) = 59.35, p < .001, \eta_p^2 = .76$, with less weight assigned to the less reliable (wide) prior overall.

Adults weighted the narrow prior more with increased experience.

This ANOVA also tested whether adults change their weight for each prior more with increasing experience (**H2**). There was a significant effect of bin, $F(9{,}162) = 5.09, p = .001, \eta_p^2 = .22$ and a significant interaction between prior and bin, $F(9{,}162) = 8.84, p < .001, \eta_p^2 = .32$. Figure 3.2 indicates that this interaction corresponds to a decreasing weight on the cue (i.e., more weight on the prior) while participants learned to use the narrow prior (blue points), but an increasing weight on the cue as participants changed their behaviour when the new wide prior was introduced (yellow points). Thus, adults showed different directions of adaptation in the two blocks (interaction), and a significant weighting change was also detected individually in the narrow prior block (Figure 3.2, blue), where weighting gradually moved towards the optimum (dashed lines). In the wide prior block (Figure 3.2, yellow), adults rapidly tended towards a ceiling effect, approaching the optimum (dashed lines). That is, they rapidly learned to mostly disregard (assign a low weight to) the new wide prior. The speed of this re-weighting from one prior to the next is considered further below.

Six- to eight-year-olds and 9-11-year-olds assigned less weight to the less reliable prior overall.

To test for age differences in overall weights assigned to priors (**H3**), we conducted a mixed ANOVA on cue weights, with age group (3 levels: 6-8-year-olds, 9-11-year-olds, and adults) as the between-subjects factor and prior variance (2 levels: narrow and wide) and bin (10) as the within-subject factors. We found a significant main effect of age ($F(2{,}48) = 7.37, p = .002, \eta_p^2 = .23$) and a significant age x prior interaction ($F(2{,}48) = 5.72, p = .006, \eta_p^2 = .19$). Simple effects analyses showed that the narrow prior was weighted more than the wide at all ages [6-8 years, $p = .01$; 9-11 years, $p < .001$; adults, $p < .001$]. This indicates that all age groups, even the youngest, changed their reliance (weighting) on the prior when its reliability changed, in line with efficient (Bayesian) decision-making that takes differing reliabilities of prior statistical distributions into account.

How rapidly did this re-weighting occur? To assess the timing of this effect, we compared the bin of 25 trials before and after the switch in prior variance using paired t-tests for each age separately (Figure 3.2, last blue vs first yellow data points). We found the wide prior weights to be significantly greater than the narrow prior weights in 9-11-year-olds ($t(16) = -3.24, p = .005$) and adults ($t(18) = -5.85, p < .001$), but not 6-8-year-olds ($t(13) = -1.13, p = .27$). These results indicate that 9-11-year-olds and adults adjust very rapidly to a change in prior reliability, while for 6-8-year-olds there is no evidence for such an immediate adjustment.

Children adapted more slowly to task statistics than adults.

This ANOVA also tested for age differences in the degree of reweighting with increasing experience (**H4**). We found an interaction between bin and age group ($F(13.215{,}432) = 1.760, p = .047, \eta_p^2 = .068$), indicating that the rates at which participants adapted to task statistics changed with age. Because adaptation for the two priors is in opposite directions (see Figure 3.2), this overall effect is difficult to interpret. To assess where adaptation was present, we ran a repeated-measures ANOVA on binned weights for each prior and age group separately. When the prior was narrow (Figure 3.2, blue points), there were no significant changes in weights across bins in 6-8-year-olds ($F(9{,}126) = .52, p = .85, \eta_p^2 = .03$), but there were in 9-11-year-olds ($F(9{,}144) = 2.31, p = .02, \eta_p^2 = .12$) and adults ($F(9{,}162) = 7.42, p < .001, \eta_p^2 = .29$). When the prior was wide (Figure 3.2, yellow points), participants of all ages increased weighting of sensory information over time [effect of bin for wide prior; 6-8-year-olds: $F(9{,}126) = 3.00, p = .003, \eta_p^2 = .17$; 9-11-year-olds: $F(9{,}144) = 2.07, p = .03, \eta_p^2 = .011$; adults: $F(9{,}162) = 5.27, p < .001, \eta_p^2 = .22$. However, Figure 3.2 shows that while adults increased their weights until they reached optimal values and remained there, 6-8-year-olds' and 9-11-year-olds' weights (top and middle panels, yellow dots) steadily rose beyond the optimal line and reached an endpoint close to a weight of 1, where there is complete reliance on sensory information.

Weighting for the cue vs. the prior is suboptimal at all ages.

Finally, we tested whether either child or adult group's weights for each prior were significantly different from optimal by the end of their experience with it (**H5**). Using one-sample t-tests, we compared each group's mean weight in the final bin of each prior to the optimal weight; see Figure 3.2, last blue and yellow point at each age vs. its corresponding dashed (optimum) line. As described in Table 3.1 and seen in Figure 3.2, by the end of their experience with the narrow (blue) prior, all age groups still placed significantly more weight on the cue than was optimal - i.e., they underweighted the prior. When the prior was wide, the weights of 9-11-year-olds and adults, but not of 6-8-year-olds, were not statistically distinguishable from optimal. Figure 3.2 indicates (last yellow points) that while adults rapidly converged on a near-optimal weight, the younger groups progressively gave (even) less weight to the prior over time, leading to a sub-optimality that, as with the narrow (blue) prior, corresponds to under-weighting of the prior – significant at 6-8 years and approaching significance ($p = .07$) at 9-11 years (Table 3.1).

Table 3.1

Results of one-sample t-tests on difference from optimal for each age group and prior.

Age group	Prior variance	t-statistic	*df*	*p* value
6-8 years	Narrow	5.76	14	<. 001
	Wide	7.75	14	< .001
9-11 years	Narrow	4.77	16	< .001
	Wide	1.91	16	.07
Adults	Narrow	4.05	18	< .001
	Wide	-1.10	18	.28

3.2.3 Discussion

In previous studies, 6-8-year-olds did not change how they weighted prior information when it changed in reliability (Chambers et al., 2018). This result was not replicated in the 6-8-year-olds in Experiment 1: children were sensitive to changes in the underlying statistical distribution. However, we found little to no effect of experience on prior weighting in 6-8-year-olds. Overall, 6–8-year-olds learned enough about the priors to treat them differently, but this learning was so slow or slight that, unlike in 9–11-year-olds and adults, it did not show measurable changes across blocks within a prior. Nine- to eleven-year-olds learned to adapt to task statistics with experience, but did so more slowly than adults. At the end of the narrow prior block, no group reached the optimum (Figure 3.2), but the 6–8-year-old group were furthest away. At the end of the wide prior block, this group were also the only one significantly different from optimal, again under-using the prior. The 6–8-year-olds were also the only group not to show a weight change from the final bin of one prior to the first bin of the next.

Why were 6–8-year-old children particularly slow to adapt to the prior? One possibility is that younger children have fundamental difficulties combining probabilistic information in a Bayesian manner because the basic mechanism required for reliability-weighted averaging is not yet developed (Chambers et al., 2018). However, children could have more basic difficulties with accurately learning the underlying statistics of the task. In our task, the ideal observer would estimate the probability distribution from all previous target locations. Working memory is needed to remember and store all target locations to successfully perform the task. However, the capacity of working memory to represent all the targets' spatial information accurately is likely to be limited – even in adults, but especially in children (Best & Miller, 2010; Diamond, 2013; Gur et al., 2012). Such limitations in capacity may mean that locations are remembered imprecisely, or that some locations, likely those seen earlier on, are forgotten with time. Experiment 2 tested the effects of working memory by asking whether children's abilities to adapt to task statistics would more closely approach those of adults if we removed the need to remember the locations of previous targets by displaying them on-screen.

3.3. Experiment 2

To test the effects of removing working memory demands, in Experiment 2, we explicitly displayed all the target locations participants had previously seen. If development of working memory is a major factor in performance, then young children in particular should show gains in adapting to task statistics as compared with Experiment 1.

We asked whether, within Experiment 2 alone, there still remained any age differences in adaptation to the task statistics (i.e., main effect or interactions including age) (**H6**). We next asked whether, comparing Experiment 2 to Experiment 1, how closely cue weights approach optimal would differ - overall, and crucially, by age group (**H7**). The display of all previous points in Experiment 2 is likely to be an advantage during the first (narrow) prior but could be a disadvantage for the second (wide), where the earlier (narrow prior) points no longer describe the current statistics. Therefore, we anticipated potentially different patterns of gain by prior (as well as by age), assessed in an experiment*age*prior interaction. Lastly, we asked whether, in this experiment, each group's performance by the end of each prior was still different to optimal (**H8**).

3.3.1 Methods

3.3.1.1 Participants. Fifty new subjects - 19 adults (mean age = 22.68 years, *SD* = 3.01), 17 6-8-year-old children (mean age = 7.51 years, *SD* = 0.92) and 14 9-11-year-old children (mean age = 10.52 years, *SD* = 0.89) - participated in Experiment 2. Three children, aged 7, 8 and 9 years, respectively were excluded from our analyses due to incomplete data. These exclusions resulted in 19 adults, 15 6-8-year-olds and 13 9-11-year-olds in the final dataset.

3.3.1.2 Task Description. The same stimuli and procedure from Experiment 1 were used in Experiment 2. However, in Experiment 2, after each trial, participants saw the target dot (feedback) for this trial (in bright red colour) as well as all target locations which participants had seen on preceding trials in transparent red colour (Figure 3.3b). These target locations remained on the screen for 1 second. Participants were told that the feedback summarized all the places the octopus had been previously. This information removed the need to memorize all the target locations and made the prior explicitly available to the participants.

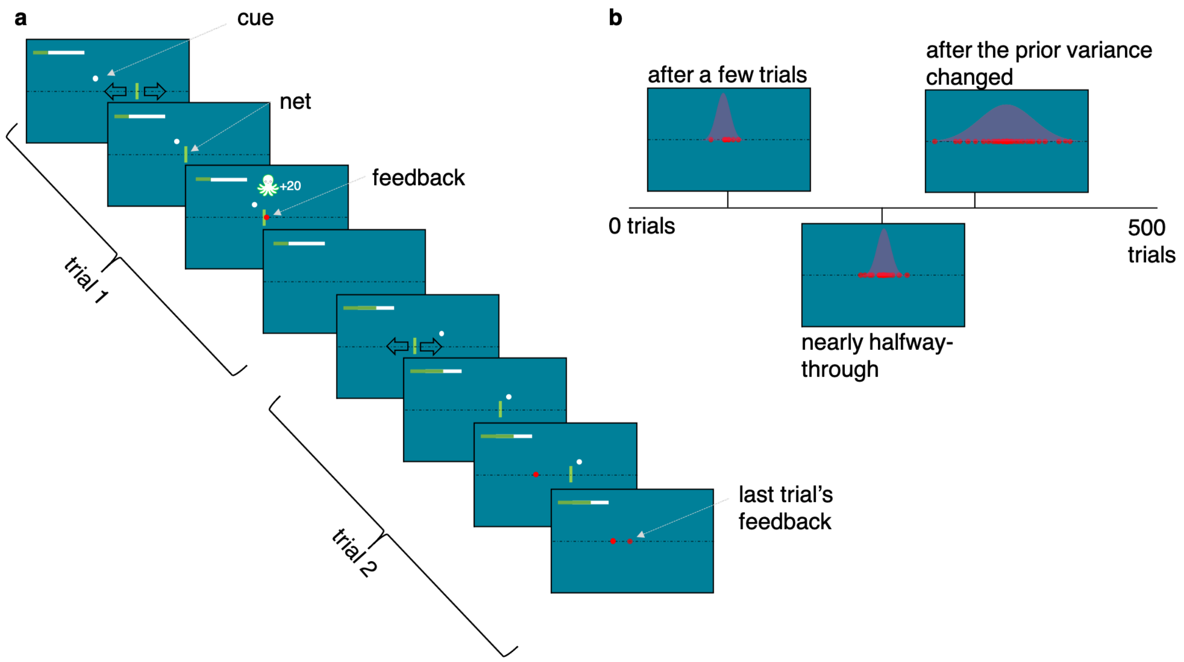

Figure 3.3. (**a**) Experimental design. The task was the same as in Experiment 1, except that as well as the feedback from the immediately preceding trial (in bright red colour), at the end of each trial, participants also saw all the feedback they had previously seen (in transparent red colour). (**b**) Timeline showing how the prior distribution unfolded over the course of the experiment, i.e., after only a few trials, nearly halfway thorough and around three-quarters of the way through, after the prior variance had been changed. For illustrative purposes, the display of all target locations (feedback) is shown with the prior distributions superimposed; participants did not see the distributions in the experiment.

3.3.2 Results

In Figure 3.4, we show cue weights replotted from Experiment 1 (solid lines) and with working memory demands reduced (Experiment 2, dotted lines). Figure 3.4 suggests that both child groups weighted the prior in a similar way to adults when demands on working memory were lower (compare square symbols across age groups), moving closer to optimal values than the children in Experiment 1 (square vs. circular symbols compared with dashed optimal prediction lines).

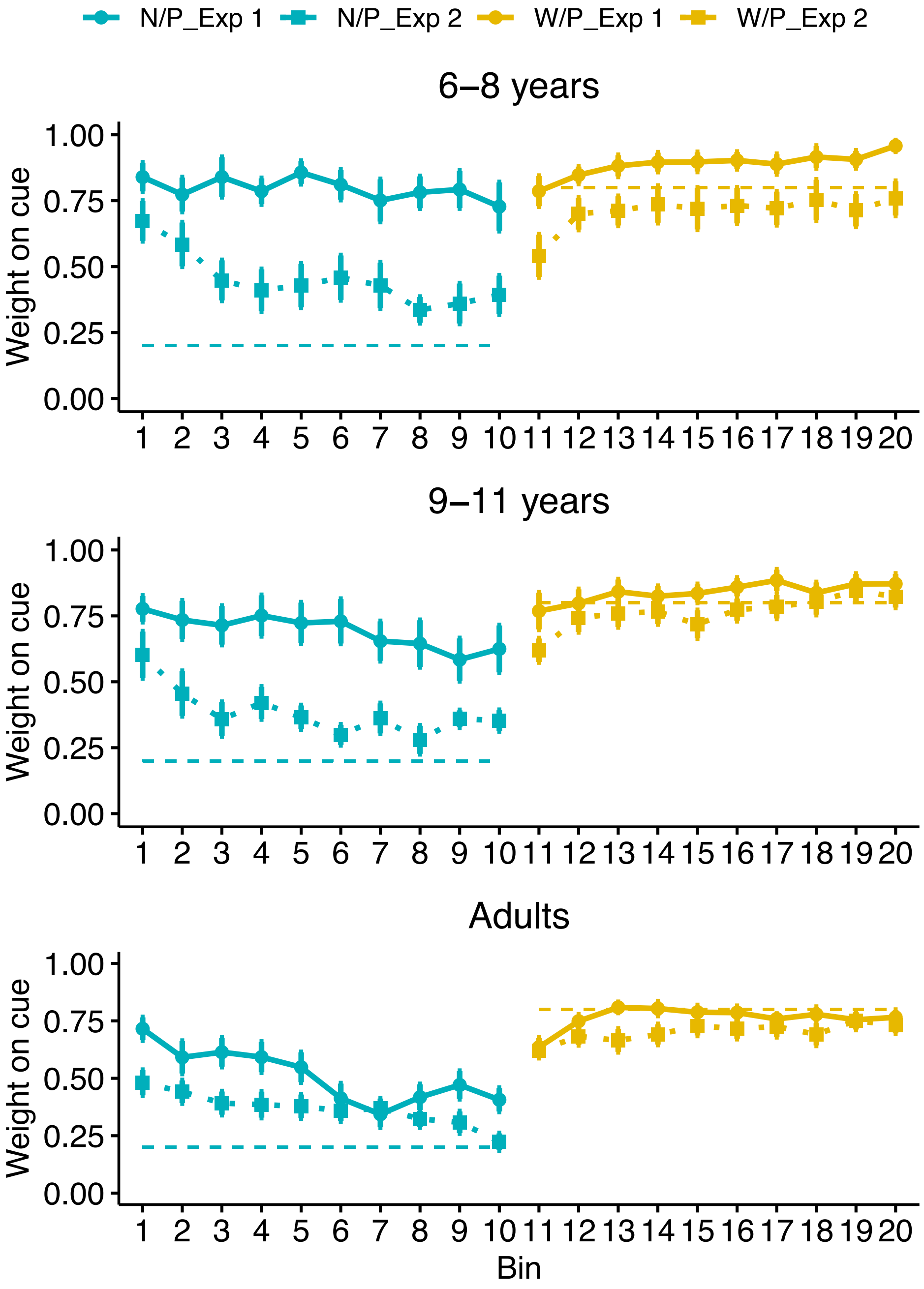

Figure 3.4. The weight given to the cue across all 20 trial bins (bin size = 25 trials) for each prior variance (blue: narrow variance, yellow: wide variance) and experiment (solid: Experiment 1, dotted: Experiment 2) for each age group (6-8-year-olds: upper panel, 9-11-year-olds: middle panel, adults: lower panel). Lower values represent a greater weight on the prior. The dashed lines indicate optimal predictions. Error bars are standard errors of the mean (s.e.m).

Adult-like adaptation to task statistics at 6-8 years after a reduction in working memory demands.

We first asked whether, when working memory demands reduced, there were still any age differences in adaptation to the task statistics (**H6**). To test this, we performed a mixed ANOVA on cue weights with prior (narrow, wide) and bins as within-subject factors and age group (6-8-year-olds, 9-11-year-olds, adults) as between-subject factors. We found a main effect of prior ($F(1,44) = 142.43, p < .001, \eta_p^2 = .76$), no main effect of bin ($F(9,396) = 1.52, p = .13, \eta_p^2 = .03$) and an interaction of prior*bin ($F(9,396) = 19.19, p < .001, \eta_p^2 = .30$), with the weight that participants placed on the narrow prior increasing as they experienced more narrow-prior trials (Figure 3.4, dotted blue lines). Critically, there was no main effect of age ($F(2,44) = .51, p = .60, \eta_p^2 = .02$) and no interaction between age and other factors (all $p > .21$). This result reflects the similarity between the dotted (Experiment 2) lines across age groups in Figure 3.4: when working memory demands were reduced, children as young as 6-8-years adapted their weights using environment statistics to a similar extent as older children and adults. The absence of an age*bin*prior interaction also suggests that the rate at which weights were adapted over the course of the session was similar across age groups.

Reduction in working memory demands affects 'distance to optimal' indices at all ages.

We asked whether how closely cue weights approach optimal differed in Experiment 2 vs. Experiment 1 (**H7**). To test this, we computed 'distance from optimal' indices by taking the absolute distance between the optimal value for each prior (0.2 for narrow and 0.8 for wide prior) and each subject's average weight in the last bin of that prior (bin 10 for narrow and bin 20 for wide prior); see Table 3.2. We then compared these indices in Experiment 2 with those in Experiment 1. A 2 (Experiment 1 or Experiment 2) x 3 (age group: 6-8 years, 9-11 years, adults) x 2 (prior: narrow or wide) mixed ANOVA on absolute distance to optimal revealed no interaction between experiment, prior and age, $F(1,92) = 1.001, p = .37, \eta_p^2 = .02$. However, we found a significant effect of prior, $F(1,92) = 39.35, p < .001, \eta_p^2 = .30$, and a significant interaction between experiment and prior, $F(1,92) = 24.68, p <$

$.001, \eta_p^2 = .21$, where absolute distance from optimal was smaller in Experiment 2 (relative to Experiment 1) when the prior was narrow ($p < .001$) but not wide ($p = .79$). We also found a significant effect of experiment, $F(1,92) = 23.93, p < .001, \eta_p^2 = .20$, significant effect of age, $F(2,92) = 9.24, p < .001, \eta_p^2 = .16$, and an interaction between experiment and age, $F(2,92) = 3.25, p = .04, \eta_p^2 = .06$. Simple effect analyses showed that 6- to- 8-year-olds ($p < .001$) and 9- to- 11-year-olds ($p < .001$), but not adults ($p = .37$), were significantly closer to optimal in Experiment 2 than in Experiment 1. Our results, therefore, showed that children showed greater gains from reducing working memory demands.

Table 3.2

Means and standard deviations (in parentheses) of absolute distance-from-optimal scores displayed for each age group, split by prior variance (narrow, wide) and experiment (1, 2).

Prior	Experiment	Age group		
		6-8 years	9-11 years	Adults
Narrow	1	.54 (.31*)*	.48 (.28)	.23 (.18)
	2	.22 (.25)	.18 (.09)	.12 (.09)
Wide	1	.16 (.06)	.14 (.08)	.09 (.09)
	2	.16 (.19)	.11 (.08)	.14 (.08)

Integration of sensory and prior information is optimal in adults but sub-optimal in children.

To test whether or not each group still differed from optimal weighting by the end of each prior (**H8**), we compared mean weights in the final bin of each prior to those predicted for an ideal observer (Figure 3.4, last blue and yellow points vs.

dashed lines). Unlike in Experiment 1, adults' weights did not differ significantly from the optimal prediction when the prior variance was narrow ($t(18) = .66, p = .51$) or wide ($t(18) = -1.97, p = .06$) – although the latter approached significance. In both child groups, weights were substantially greater than the optimal predictions when the prior variance was narrow [6-8 years: $t(14) = 2.64, p = .01$; 9-11 years: $t(12) = 3.89, p = .002$]. However, when the prior variance was wide, performance was indistinguishable from optimal for both 6-8-year-olds ($t(14) = -.63, p = .53$) and 9-11-year-olds ($t(12) = .55, p = .59$).

3.3.3 Discussion

Six- to eight-year-olds in Experiment 1 distinguished between the two priors, but were slow to adapt to each distribution, not showing weighting changes across bins, and remaining far from optimal. Strikingly, the rate and extent of their adaptation became adult-like when we made all previous trial outcomes explicitly available in Experiment 2. Together, these results suggest that younger children are capable of adapting to task statistics in an adult-like manner when those statistics need not be represented in working memory. Therefore, memory limitations are likely to be a major factor contributing to young children's slower adaptation to task statistics.

Having identified memory as a challenge for young children's perceptual decision-making in novel environments, we turn to another: the need, potentially, to deal with multiple uncertainty levels of either prior or sensory (likelihood) information. While 6-8-year-olds in Experiment 1 were able to distinguish the two levels of prior uncertainty, 6-8-year-olds in Chambers et al.'s (2018) study could not. However, Chambers et al. (2018) had three levels of sensory uncertainty that changed from trial to trial, whereas the sensory uncertainty in Experiment 1 was constant throughout the experiment. Could it be that 6-8-year-olds in Experiment 1 had an easier time differentiating between the prior variances because they did not have to learn and apply multiple weighting rules simultaneously? Application of multiple rules and switching between them from trial-to-trial, places greater demands on executive function and cognitive flexibility, developing substantially throughout childhood

(Carlson et al., 2013). We test the effects of increasing these demands in Experiment 3.

3.4 Experiment 3

We asked how children’s abilities to differentiate between prior variances are affected by adding a second level of sensory uncertainty. We looked for differences in weights given to different priors and likelihoods, and age differences in these, with a specific interest in whether each age group would still differentiate between the prior variances, giving less weight to the less reliable prior (**H9**). We also asked whether, given the same number of trials per prior, but having to deal with multiple likelihoods, adults or children would be less optimal than in Experiment 1 (**H10**).

3.4.1 Methods

3.4.1.1 Participants. Eighteen adults (mean age = 20.27, age range 18-25 years of age), 15 6-8-year-old children (mean age = mean age = 7.61, *SD* = 0.99) and 17 9-11-year-old children (mean age = 10.44, *SD* = 1.11) took part.

3.4.1.2 Experimental Procedure. The basic task parameters were otherwise similar to those of Experiment 1, except that we included an additional level of sensory uncertainty. As in Experiment 1, the horizontal position of the target was sampled from a narrow Gaussian distribution in the first half of trials (trial 1-250) and a broad distribution in later trials (trial 251-500). We used the same variance of the likelihood ($\sigma_L = 0.1$) as in Experiment 1 but also added a second, low-noise, likelihood variance ($\sigma_{Ll} = 0.02$). In each trial, we pseudo-randomly chose one of the two variances of the likelihood while ensuring an equal number of trials for each likelihood and drew the cue dot by sampling from a normal distribution with the chosen variance. We therefore tested four conditions in Experiment 3: low-noise and high-noise likelihood paired with narrow and wide prior variances. Subjects completed one 1-hr session and experienced 250 trials per prior variance (500 trials overall), as before, except that they were split between likelihood variances.

3.4.1.3 Data Analysis. To more directly compare prior weightings across Experiments 1 and 3 (where there were the same number of prior trials, but these were split between two likelihood variances), for every participant, we extracted the trials separately for each prior variance. We binned the data for each prior into 10 bins (bin size = 25 trials); the same as Experiment 1. In each bin, we looked at trials with each likelihood variance separately. The number of trials for a given likelihood variance assigned to each bin differed – there were at least 11-16 trials of each likelihood variance assigned to each bin. As described in Experiment 1, we computed the relative cue weights for each of the 20 bins (10 for each prior variance) by regressing each participant's response against the location of the sensory cue.

3.4.2 Results

Cue weights replotted from Experiment 1 (solid lines) and Experiment 3 (dotted lines) are shown in Figure 3.5. Figure 3.5 suggests that in Experiment 3, child and adult groups performed similarly (compare square symbols across groups), with adults weighting the prior less than in Experiment 1 (compare same colour conditions, with equal reliabilities, for Experiment 3 – square vs. Experiment 1 – circle).

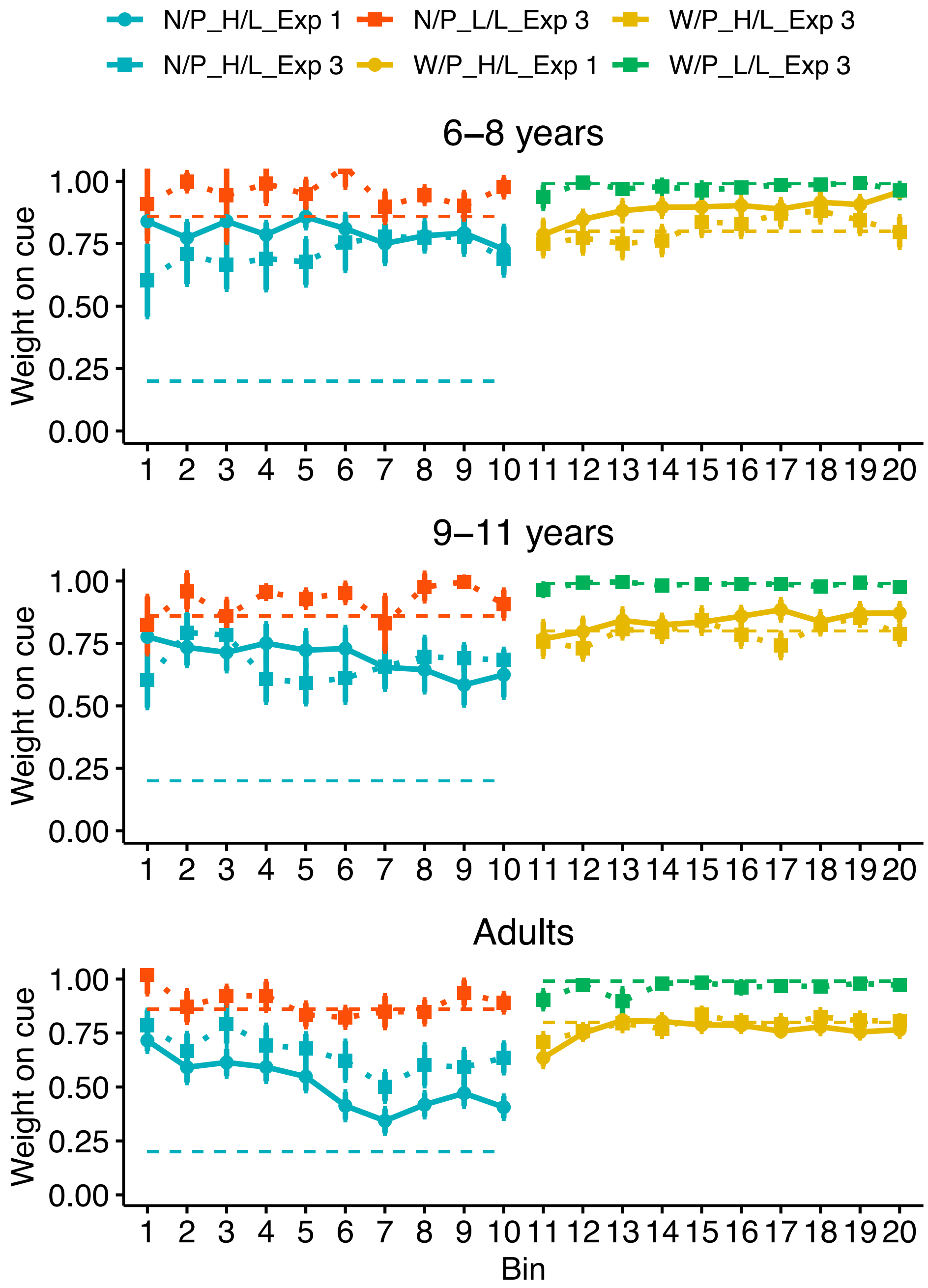

Figure 3.5. Cue weights across 20 trial bins for each prior and likelihood pairing (blue: narrow prior, high likelihood (N/P_H/L), red: narrow prior, low likelihood (N/P_L/L), yellow: wide prior, high likelihood (W/P_H/L), green: wide prior, low likelihood (W/P_L/L) and age group (top: 6-8 years, middle: 9-11 years, bottom: adults) for Experiment 3 (filled squares, dotted lines). The data for the prior/ likelihood pairings which were the same in Experiment 1 are also replotted (filled circles, solid lines) for comparison. Lower values represent a greater weight on the

prior. The horizontal dashed lines indicate optimal predictions. Error bars are standard errors of the mean (s.e.m).

Six- to- eight-year-olds do not distinguish between levels of prior uncertainty when multiple interleaved levels of sensory uncertainty were introduced.

A mixed ANOVA with within-subjects factors prior variance (narrow, wide), likelihood variance (low, high) and bin, and between-subjects factor age group (6-8 years, 9-11 years, adults) revealed a main effect of prior ($F(1,47) = 23.39, p < .001, \eta_p^2 = .33$), with weights being significantly lower for narrow variance relative to wider variance (Figure 3.5, dotted lines). There was no interaction between prior and age ($F(2,47) = .55, p = .58, \eta_p^2 = .02$), suggesting that the degree of reweighting across priors did not vary as a function of age. We also found a main effect of likelihood ($F(1,47) = 99.86, p < .001, \eta_p^2 = .68$) but no interaction between likelihood and age ($F(2,47) = .70, p = .93, \eta_p^2 = .003$), indicating that younger children use sensory information to a similar extent as older children and adults. There were no other significant interactions involving the factor age, and no main effect of age (all $p > .13$). Thus, in contrast to Experiment 1 (Figure 3.5, solid lines), which showed strong age differences, in Experiment 3 (Figure 3.5, dotted lines), performance across age groups was similar. Overall, participants were sensitive to the new addition – trial-to-trial changes in the likelihood, but there were no age changes in this sensitivity.

Simple effects analyses tested whether each age group differentiated between the two priors (**H9**). These revealed that when there were multiple interleaved levels of sensory uncertainty, 9-11-year-olds and adults successfully differentiated more reliable from less reliable priors ($p = .004$ and $p = .01$, respectively); 6-8-year-olds, however, did not ($p = .07$), suggesting that younger children struggle to re-adjust to changes in the environment when they had to learn multiple weighting rules at the same time.

Participants were not less optimal when dealing with multiple likelihoods.

As in Experiment 2, we examined whether performance in Experiment 3 would be more or less optimal than performance in Experiment 1. Specifically, we tested whether, given the same number of trials per prior, but having to deal with multiple likelihoods, subjects would be less optimal (**H10**). As in Experiment 2, for each prior variance separately, we calculated the absolute distance between subjects' average weights in the last bin of the prior and the optimal value for that prior (Table 3.3). We analysed only the trials in Experiment 3 with the high-noise likelihood that is the same across experiments 1 and 3. The resulting values were then submitted to a 2 (prior: narrow, wide) x 2 (experiment: Experiment 1, Experiment 3) x 3 (age group: 6-8-years, 9-11-years, adults) ANOVA. We observed a main effect of prior ($F(1{,}95) = 139.14, p < .001, \eta_p^2 = .59$), a main effect of age ($F(2{,}95) = 8.72, p < .001, \eta_p^2 = .15$), but no main effect of experiment ($F(1{,}95) = .04, p = .83, \eta_p^2 < .001$). There was also no significant age * experiment interaction ($F(2{,}95) = 1.78, p = .17, \eta_p^2 = .03$), suggesting that none of the age groups moved further from optimal values when they had to alternate between two likelihoods compared to when there was only one likelihood (see also Table 3.3).

Table 3.3

Means and standard deviations (in parentheses) of absolute distance-from-optimal scores displayed for each age group, split by prior variance (narrow, wide) and experiment (1, 3).

Prior	Experiment	Age group		
		6-8 years	9-11 years	Adults
Narrow	1	.54 (.31*)*	.48 (.28)	.23 (.18)
	3	.49 (.23)	.48 (.16)	.43 (.29)
Wide	1	.16 (.06)	.14 (.08)	.09 (.09)
	3	.17 (.15)	.12 (.09)	.08 (.04)

3.5 General Discussion

The present experiments asked why young children might adapt to novel task statistics less efficiently than older children and adults. Using a spatial localisation task, Experiment 1 found that children aged 6-8 years were sensitive to underlying task statistics, differentiating between two statistical prior distributions. However, they adapted to these more slowly and less completely than adults did. Nine- to eleven-year-olds adapted with experience but more slowly than adults. We next probed two potential developmental factors limiting young children's efficient use of task statistics: (i) working memory limits (Experiment 2), and (ii) cognitive flexibility; specifically, difficulty in managing multiple rules (Experiment 3). We will discuss each in turn.

Removing the need to learn task statistics and represent these in memory (by making previous target locations explicitly available) in Experiment 2 brought 6-8-year-olds' weights close to those of older children and adults. Strikingly, with the memory demands removed, there were no significant differences in performance across age groups, although children's weighting at the end of the first prior still differed significantly from optimal. Our finding of greater, much more adult-like, adaptation to statistics in 6-8-year-olds when the demands on memory were reduced is in line with marked improvements in working memory during childhood (Zelazo et al., 2008). This result goes against the idea that children cannot optimally combine before the age of 9 mainly because mechanisms for weighted information integration are undeveloped (Chambers et al., 2018; Ernst, 2008). Instead, our data appear to be more consistent with the emerging idea that children are fundamentally limited by 'resource constraints', which can arise from limited cognitive functions, such as working memory - a conclusion compatible with the 'resource rational' account recently proposed by Lieder and Griffiths (2019). It is possible that the reason why the 6-8-year-olds in Experiment 1 did not adapt to task statistics over time was because their working memory capacity did not allow them to retain accurately the preceding target locations needed to correctly infer the underlying structure of the task.

Interestingly, our results suggest that central cognitive resources also limit adult performance, albeit less drastically: reducing working memory demands improved performance across all age groups, including adults. When the need to remember previous target locations was removed in Experiment 2, adult performance was indistinguishable from optimal, unlike in Experiment 1. Sub-optimal use of prior information in some previous studies with adults (e.g., Bejjanki et al., 2016; Vilares et al., 2012) may be explained in part by limitations in even adults' abilities to remember and represent all previous trial outcomes. We note that learning the prior distribution could be accomplished equivalently by storing all trials in memory, or (computationally less demandingly) by updating summary statistics for a distribution based on appropriately weighted new samples. These processes can be difficult to distinguish (Hansmann-Roth et al., 2021; Spicer et al., 2020; Sun et al., 2019), but both require accurate representation and updating in working memory.

We further showed that while 6-8-year-old children distinguished between two differently reliable priors when they had to learn only one weighting 'rule' for each (Experiment 1), they were no longer able to do so when they had to learn multiple weighting rules simultaneously, to deal with two different levels of sensory noise (Experiment 3). This result is consistent with Chambers et al. (2018), who also found this lack of prior weighting in a situation with multiple sensory likelihoods. These results are in line with extended development of cognitive flexibility: young children have shown difficulty in managing multiple rules (Anderson et al., 2011; Harada et al., 2018; Huizinga & Van Der Molen, 2007; Zelazo et al., 1996, 2008). Some studies which have a multiple-rules paradigm showed age-related improvements in performance between the ages of 5 and 8, with adult-like performance in children older than 8 years of age (Irwin-Chase & Burns, 2000). In two recent experiments closely related to the present question, Bejjanki et al. (2019) showed that 6-7-year-old children changed their weightings when only the sensory uncertainty changed (i.e., single 'rule), but not when both sensory and prior uncertainty changed in an interleaved fashion (multiple 'rules' at the same time). These findings support the idea that having to store and manage several weighting 'rules' may have made 6-8-year-olds less sensitive to changes in the underlying distributions. Interestingly, this manipulation, too, had a major impact on adults' performance. Adult performance now became so 'child-like' that there were no longer any significant age differences.

Taken together, our results suggest that although children are sensitive to statistical regularities in their environment from an early age, their abilities to use these to make effective perceptual decisions depend on development of later-developing central cognitive abilities, including working memory and cognitive flexibility. The suggested memory limitations would imply that when children or infants have been shown to detect statistical patterns, the fidelity with which these have been learned may improve markedly with age. Actually determining what has been learned remains a challenge for either studies with infants (where just measuring sensitivity vs chance does not provide this information), or older children (where, as here, measuring behavioural choices does not separately measure what was learned vs how it was acted upon). New approaches to inferring the statistics that are represented – using proxy measures such as responses to unexpected stimuli from pupillometry or EEG – may be important future developments in this field. Other important directions include testing the degree to which either individual differences in, or direct manipulations to, these core cognitive skills, can predict efficient use of statistics in children or adults.

We conclude that accounting for cognitive limitations during development is essential for understanding the mechanisms underlying children's – and adults' - abilities to perceive and act efficiently when faced with novel environments. In the real world, such learning is needed in new visuo-motor tasks, new real-world spaces, and potentially more abstractly, in new social or educational settings. Future research to optimise abilities to learn, accounting for developmental cognitive limitations, has potential applications to atypical development, disorders, and optimal approaches to education and training in childhood and beyond.

3.6 From Experimental Manipulations to Individual Differences

Across three experiments, we explored several lines of evidence for which factors may affect the learning and efficient use of novel task statistics during development. We found that removing the need to learn the statistics and represent these in memory (by making previous target locations explicitly available) in

Experiment 2 brought 6-8-year-olds' weights close to those of older children and adults, although they still remained far from optimal. Conversely, when they had to learn multiple weighting rules simultaneously (increased cognitive effort) in Experiment 3, younger children's abilities to distinguish between the priors broke down. Given the seemingly critical role of working memory capacity and cognitive flexibility in the efficient adaptation to task statistics, and the considerable individual differences we observed, it is possible that individual differences in these factors may also predict performance in adults. In Chapter 4, we focus on individual differences in working memory.

Chapter 4

Individual differences in working memory capacity do not predict efficient adaptation to novel statistics

4.1 Introduction

Working memory has a central role in our ability to uncover statistical regularities in our environment (Baddeley, 2003, 2007; Bianco et al., 2020). As was argued by Bianco et al. (2020), detecting these regularities requires each event that is part of this sequence (in our case corresponding to the history of all previous target locations) to be retained in memory for long enough for the observer to figure out how they fit together. Although Bayesian models assume that behaviour is equally affected by all previous events (Geisler, 2011; Kersten & Yuille, 2003), it is already well known that while we can probably remember what happened on the immediately preceding trial with relative ease, we may not be able to remember what happened ten or twenty trials ago; certainly not with the same quality of representation as one trial back (Cashdollar et al., 2017; Luck & Vogel, 2013; Ma et al., 2014). In keeping with this notion, several studies have shown that their data are well explained by fitting an exponential function that weights more recent trials more than trials experienced earlier on (Bornstein & Daw, 2012; Harrison et al., 2011). It follows that if our participants are not able to retain in memory all the feedback they received so far or can only recall up to several trials into the past, they will not be able to infer the correct underlying statistical regularities in their environment. Given the limits on the capacity of working memory, we expected that higher working memory capacity, permitting the storage of more information, would be associated with more efficient adaptation to novel task statistics (i.e., be closer to optimal values). Indeed, when we made all past target locations explicit in Experiment 2, Chapter 3, freeing up memory capacity, both children and adults moved closer to optimal values, with adult performance indistinguishable from optimal when the prior was narrow. Higher working memory capacity has also been linked to better discrimination and faster responsiveness in alternative forced-choice tasks (Ester et al., 2014) and better implicit learning of sequences (Frensch & Miner, 1994; Usher & McClelland, 2001; Virag et al., 2015). Shared neural underpinnings between the storage capacity of working memory and decision making have also been reported (Morcos & Harvey, 2016; Romo et al., 1999; Schott et al., 2019; Shadlen & Newsome, 2001).

To this end, we were interested in exploring whether differences in working memory capacity explain inter-individual variations in how close adult participants are to optimal. We used the same task that was used in Chapter 3, in which participants had to localise unseen targets using a noisy sensory cue and a prior distribution that can be learned over the course of the experiment. We decided on distance-to-optimal score as our outcome of interest because we were interested in whether higher working memory capacity is associated with more efficient adaptation to the underlying task statistics.

Due to the spatial nature of our task, in the present study, we focussed on the spatial component of working memory (Baddeley, 2003). Only one test, Corsi block task, directly assesses the capability to remember spatial positions and the sequence in which they occur (Corsi, 1972; Della Sala et al., 1999; Kessels et al., 2000). Here, two blocks light up one after the other, and participants have to tap the blocks in the same order, with the number of blocks increasing until performance breaks down (Corsi, 1972; Kessels et al., 2000). We had decided on the Corsi block test because efficient adaptation to statistics of spatial sequences depends on abilities of spatial memory and sequential structure. Spatial memory is crucial because inferring the probability distribution over target locations (i.e., the prior) would require remembering all previous target locations. Holding the sequence of trial outcomes in memory – how far back in time a given location appeared (e.g., previous trial vs. ten trials in the past) – is important because in the environment that change, more recent experiences might signal changes in the environment and should therefore be favoured over earlier experiences. As a consequence, it is necessary to test whether observers remember spatial sequences. For this reason, our study included tasks in which subjects had to recall spatial positions and sequences.

Alternatively, we considered the possibility that observers use different strategies to better remember these positions. For example, one observer may visualize the spatial environment, while another may internally verbalize and rehearse the spatial locations they were seeing (e.g., 'the target appeared slightly to the right of the centre of the screen'). We reasoned that if verbal memory plays a role in the acquisition of statistical regularities, we should find a negative correlation between verbal working memory ability and distance-from-optimal score from the localisation task. For this purpose, we make use of the backwards digit task (Waters & Caplan, 2003; Wilde et al., 2004), which is one of the most established verbal working memory span tasks (Conway et al., 2005). In addition, to avoid a confound of general cognitive ability, we included a measure of vocabulary knowledge measured by the vocabulary sub-test from the Wechsler Abbreviated Scales of Intelligence - 2nd edition (WASI-II) (Wechsler, 2011) as a predictor.

To explore the extent to which efficient use of task statistics in adults depends on working memory, a multiple linear regression was conducted. If we find that visuospatial, and perhaps verbal working memory significantly predict distance from optimal, we can conclude that individual differences in working memory explain a significant amount of variance in how efficiently different adults use task statistics, above and beyond general cognitive ability. On the other hand, if we find that neither of the working memory measures makes a significant contribution, we can conclude that working memory is not a major source of variation. We were primarily interested in the role of visuospatial working memory – therefore, we conducted a hierarchical regression analysis which allowed us to determine whether visuospatial working memory contributed unique, independent variance beyond that explained by vocabulary knowledge (added in Step 1) and verbal working memory (added to vocabulary score in Step 2).

4.2 Materials and methods

4.2.1 Participants

Sixty-three adults ranging in age between 18 and 35 years with normal or corrected-to-normal visual acuity participated in the experiment voluntarily. One participant was excluded from these analyses due to incomplete data. The remaining 62 participants (mean = 20.93; SD = 1.96) were included in our analyses.

The sample size of $N = 64$ was selected so that in the multiple regression analysis containing three predictors the power to detect a moderate deviation of a single linear regression coefficient from 0 (partial R^2 = .15) was 1 – β = 0.8 (two-tailed test), with α set to 0.05. According to G*Power, the required minimum sample size is 64.

Participants were recruited from the university's subject pool and received course credit or were paid £15 for their participation. The experiment was conducted according to the principles expressed in the Declaration of Helsinki. All adults participated voluntarily after providing informed written consent. The study was approved by the ethics committee of the Department of Psychology, Durham University.

4.2.2 Measures

4.2.2.1 Experimental Task. The apparatus, stimuli and task were the same as described in Experiment 1, Chapter 3. Participants completed a spatial localisation task, where the target was more likely to be in the middle of the screen, either clustered closely (narrow variance) or spread out over a larger region (i.e., wide variance). A single dot stimulus drawn from a Gaussian distribution gave participants noisy information about the current location of the target, and participants were asked to estimate its location. The performance measure was a distance-from-optimal score.

4.2.2.2 Measures of Working Memory. The Psychology Experiment Building Language program (PEBL; Mueller & Piper, 2014) was used to run the working memory tasks. The Corsi Test (visuospatial working memory) and Backwards Digit span (verbal working memory) were the measures of working memory, and these have been validated in adults (Kessels et al., 2000).

Corsi Test

The Corsi Test (Corsi, 1972) provided a measure of spatial working memory. For this test, participants saw a display of nine blue blocks on the screen. The blocks lit up (changed colour to yellow) in a particular order, and the participant used the mouse to click on the blocks in the correct order. Feedback was provided as to whether the participant was correct or incorrect. First, the participant did 3 practice trials of three blocks. They then started with 2 blocks to remember and had 2 trials at that level. If they got at least one right, the sequence that must be remembered increased by one. If they missed both trials at the same sequence length, the program stopped. Memory span was computed as the smallest number of blocks (2) plus the total number correct and divided by the number of blocks at each sequence length.

Backwards Digit Span

This task provided a measure of verbal working memory. Participants were shown a string of digits (e.g., 376) in the middle of the screen. The digits were presented sequentially for 1.5 seconds per digit. Following presentation of the string of digits in random order, participants were instructed to use the numbers on the keyboard to type the digits in reverse order (in this case, 673). Feedback of “correct” or “incorrect” was provided, as well as the correct answer. The string increased by one digit every two trials, from a two-digit string to a maximum of nine digits. The task was stopped after two successive incorrect responses on the same digit string. Digit span was computed as the maximum number of digits participants could repeat in reverse order on at least one trial.

Vocabulary

The Vocabulary subtest from the Wechsler Abbreviated Scale of Intelligence (WASI-II) was used to assess participants’ vocabulary knowledge. In this test, the experimenter read out a series of words from the WASI II booklet and the participant was asked to describe what they mean. The experimenter then wrote down the response in the answer booklet. The responses were manually scored using the WASI-II manual after the testing session. There were 31 words in total (3 picture and 28 verbal), with a maximum score of 59. Testing was terminated after 3 consecutive scores of 0. The total correct was calculated by summing the total correct responses (2 = correct, 1 = partially correct, 0 = incorrect).

4.2.3 General Procedure

The task order was pseudo-randomised across participants. A list of all 24 possible sequences was created, and each new participant was assigned the next available sequence down the list. After all the sequences were completed once, the same sequences were assigned again. The entire testing lasted approximately 1.5 hours.

4.2.4 Data Analysis

Data processing, analysis and ideal observer calculations are the same as those described in Chapter 3 (details on pages 92-93).

Individual differences were analysed via a multiple linear regression. Our outcome measure, distance-from-optimal score, was computed by subtracting the optimal values (0.2 for narrow prior; 0.8 for wide prior) from each subject's weight averaged across the last 25 trials of each prior (details in Chapter 3). We included three predictors: (1) the memory span in the Corsi-blocks task that indexes visuospatial working memory capacity, (2) the digit span in the Backwards Digit span task that indexes verbal working memory capacity, and (3) the number of correctly described words in the vocabulary subtest from the WASI-II, which was included as a measure of vocabulary knowledge (for a description of how these measures were calculated, see Materials and methods).

We used hierarchical regression to assess the relative contributions of the different predictor measures to distance-from-optimal indices. We used F ratio statistics to determine the amount that the explained variance changed from one model to the next.

4.3 Results

We begin by repeating the same analysis used to analyse the data from the adult participants in Chapter 3 (rm-ANOVA on cue weights with prior (narrow, wide) and bin as factors) to ensure that we could replicate their pattern of performance in this new sample of adult participants.

Adults assigned less weight to the less reliable prior overall.

We conducted a 2 (prior variance: narrow vs. wide) x 10 (bin) repeated-measures ANOVA on the cue weights. We found, similarly to our findings in Chapter 3, a main effect of prior, $F(1,62) = 117.86, p < .001, \eta_p^2 = .65$, with less weight assigned to the less reliable (wide) prior overall.

Adults weighted the narrow prior more with increased experience.

This ANOVA also showed, similarly to our findings in Chapter 3, that there was a significant effect of bin, $F(6.38, 395.83) = 3.19, p = .004, \eta_p^2 = .04$, and a significant interaction between prior and bin, $F(5.74, 356.37) = 13.993, p < .001, \eta_p^2 = .18$. Figure 4.1 (blue solid line) and simple effects analyses show that the weight given to the cue decreased over time when the prior was narrow; the first significant drop in cue weight was in bin 2 ($p = .001$), and the weight dropped even further in bins 8-10 (all $p < .01$). In the wide prior block, the weight placed on the cue increased significantly from bin 11 until reaching a plateau around bin 12 ($p < .001$), Figure 4.1 – compare yellow solid and dashed lines.

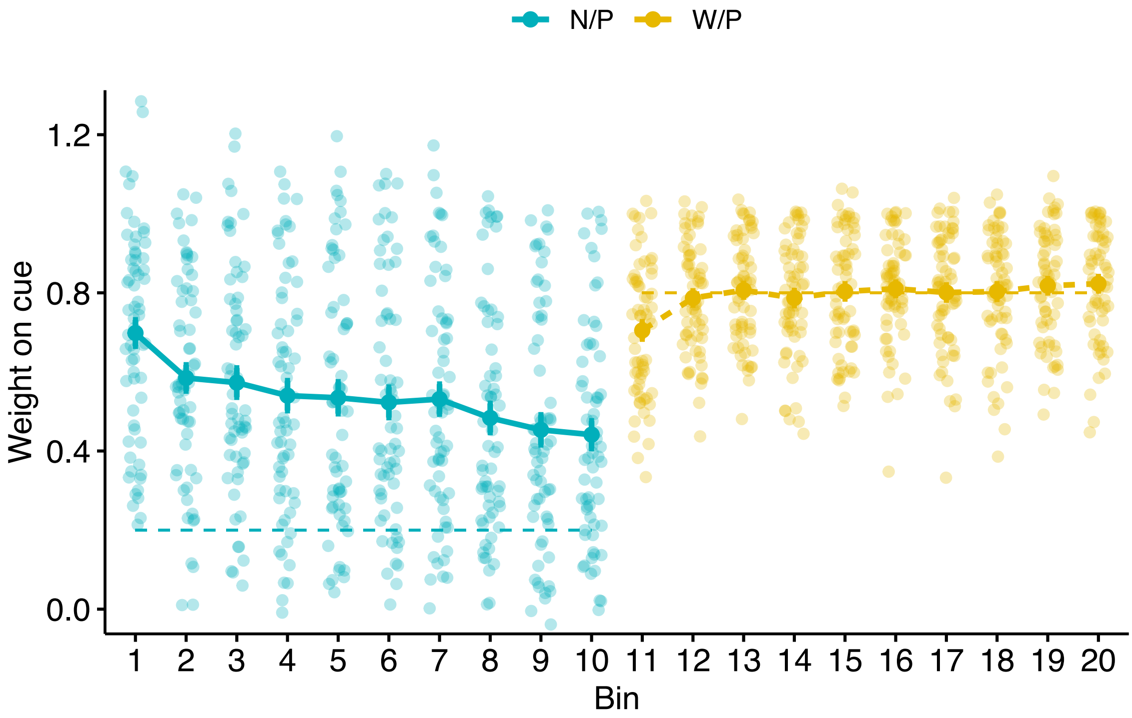

Figure 4.1. Mean weight placed on the cue, separated by bin and prior variance (blue: narrow variance, yellow: wide variance). There are 25 trials in each bin. Lower

values represent a greater weight on the prior. Each circle is a participant, and error bars are standard errors of the mean (s.e.m). The dashed lines indicate optimal predictions.

Altogether, we found a similar pattern of results as those reported in Chapter 3. Here, it is important to highlight that, similarly to the adults in Experiment 1, Chapter 3, adults in our current study rapidly approached the optimum when the prior was wide (compare solid yellow line vs. dashed yellow (optimal) line, Figure 4.1). By contrast, although observers moved gradually towards the optimum, they were still far from it by the end of the narrow prior trials (solid blue line vs. dashed blue (optimal) line, Figure 4.1), showing that adult observers, overall, did deviate from optimal. As Figure 4.1 illustrates, there was also a spread in these deviations from optimal between individuals (compare light blue and yellow dots vs. their corresponding solid line), suggesting that there are considerable individual differences.

Regression results

We used three measures as predictors: 1) Corsi memory span – calculated by adding the smallest number of blocks that needed to be remembered in the correct sequence and the total number correctly remembered, and then dividing that by the number of blocks at each sequence length; 2) Backwards digit span – computed as the number of digits correctly repeated in reverse order on at least one trial, and 3) vocabulary score: total number correct. Frequency distributions for all three measures are shown in Figure 4.2 (A: Corsi, B: Backwards digit, C: vocabulary). The skewness for the distributions is small and clustered around the mean (red line) as expected for normal distributions. From these histograms we conclude that the predictors are approximately normally distributed, with the mean (red vertical line) and variance within the expected range (Kessels et al., 2000; Wechsler, 2011)

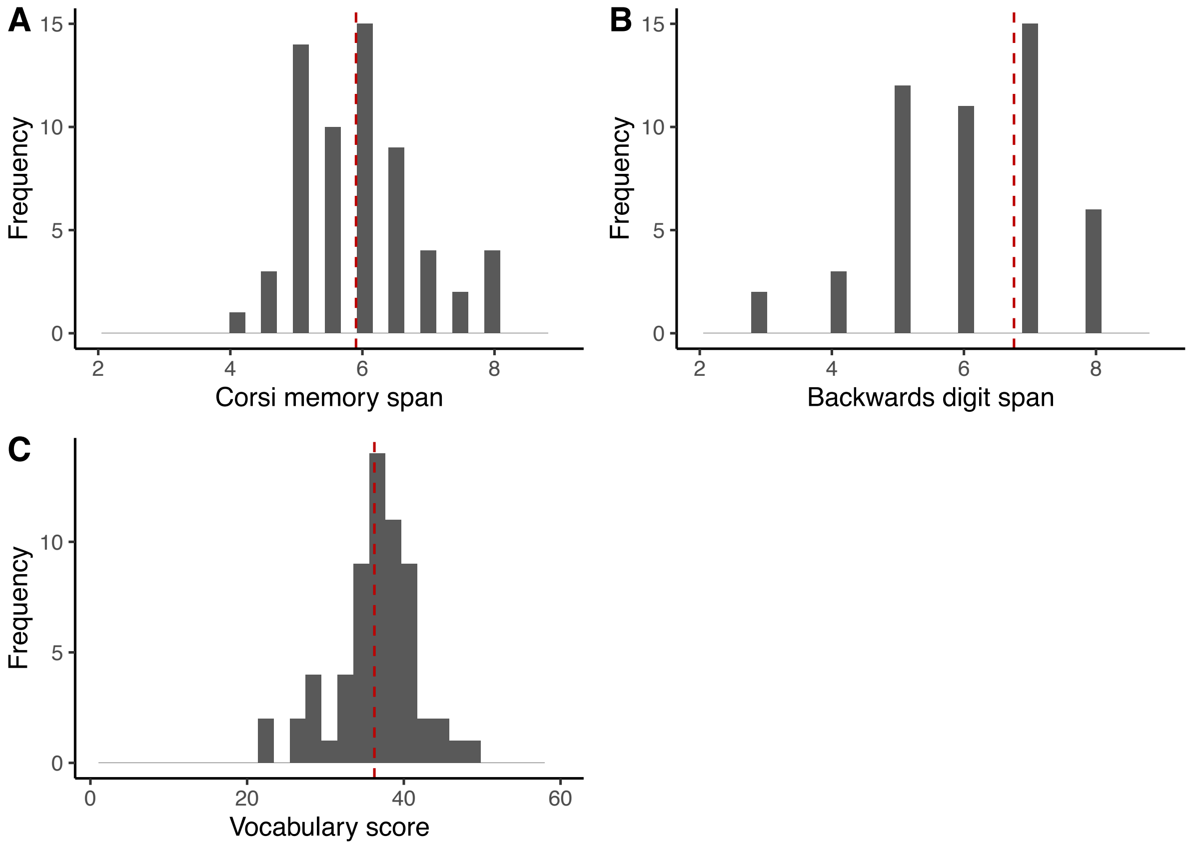

Figure 4.2. Histograms showing the distribution of scores for the Corsi blocks test **(A)**, Backwards digit span **(B)** and vocabulary **(C)**. Red vertical lines represent the mean.

However, our outcome variable (distance-to-optimal scores) followed a right-skewed distribution both when the prior was narrow (Figure 4.3A) and wide (Figure 4.3B), and more importantly, these distributions were significantly different ($D = .41, p < .001$, two sample Kolmogorov-Smirnov test). Given that the two priors (narrow, wide) showed differences in variability, this would have violated the homogeneity of variance assumption for the regression analysis reported below. Such a violation is known to result in a higher Type I error rate. Therefore, the regression analysis is conducted separately for the narrow prior and the wide prior.

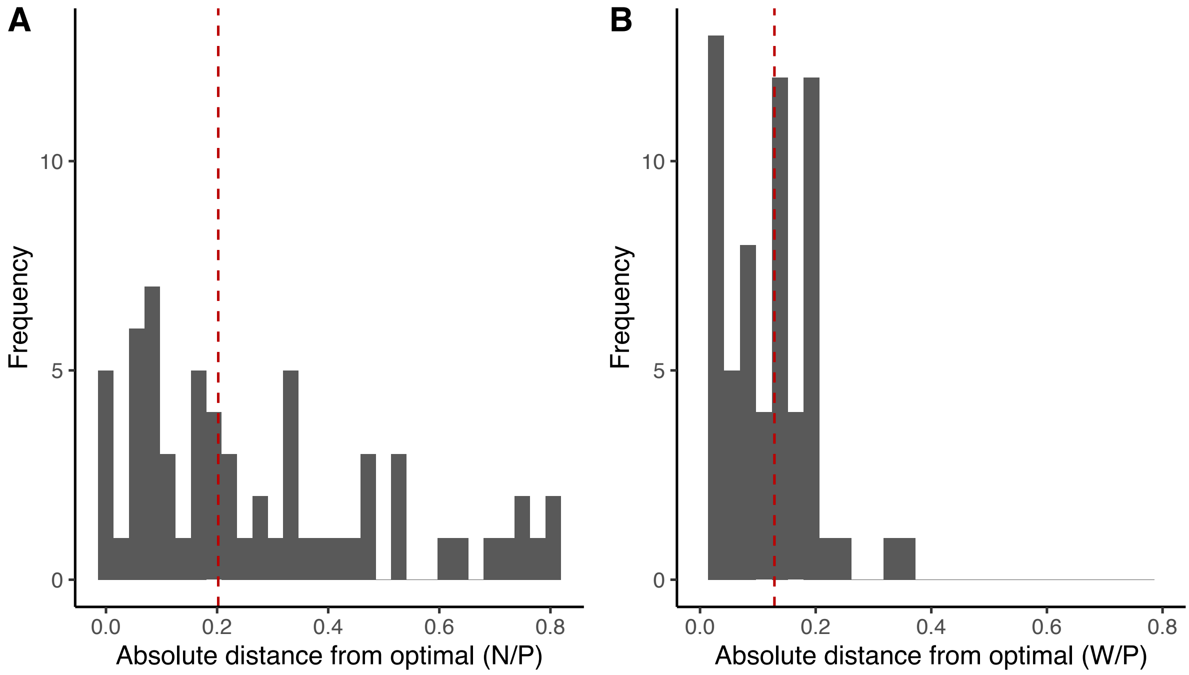

Figure 4.3. Histograms showing the distribution of (absolute) distance-to-optimal indices when the prior was narrow **(A)** and wide **(B)**. Red vertical lines represent the median.

Critically, we observed considerable variability in performance across individuals in our outcome and each of our predictor measures (Figures 4.2 and 4.3 and Table 4.1).

Table 4.1

Shown are the means and standard deviations (SD; in brackets) and range for both outcome measures (distance-from-optimal for narrow and wide prior) and predictor measures (vocabulary, Backwards digit span, Corsi blocks)

Measure	Mean (SD)	Range
Distance-from-optimal (narrow prior)	0.28 (0.23)	0.0-0.8
Distance-from-optimal (wide prior)	0.11 (0.07)	0.01-0.35
Corsi blocks (/9)	5.90 (0.93)	4-8
Backwards digit span (/10)	6.75 (1.79)	3-10
Vocabulary (/59)	36.24 (5.00)	23-48

None of the predictors made unique contributions to variability in distance-to-optimal scores.

We next sought to determine whether each of our predictor measures (vocabulary score, backwards digit span, Corsi memory span) explain unique variance in distance-to-optimal indices by using a hierarchical regression (see Table 4.2 for regression parameters). The residuals were not normally distributed. Therefore, we transformed the data by taking the square root of the distance-to-optimal scores and fitted the models once again. By comparing nested linear models, we found that compared to a model with vocabulary alone (adjusted $R^2 = -.01$), adding backwards digit span did not explain significant additional variance in distance-to-optimal scores (model comparison: narrow prior: $F(2,59) = .19, p = .82$, adjusted $R^2 = -.02$; wide prior: $F(2,59) = .30, p = .73$, adjusted $R^2 = -.02$. Moreover, addition of Corsi memory span explained a further 2.6% variance in performance when the prior was narrow and 0.6% when the prior was wide, but this was not significant for either prior (model comparison: narrow prior: $F(3,58) = 1.55, p = .21$; wide prior: $F(3,58) = .32, p = .80$).

Table 4.2

Shown are the unstandardized regression coefficients (β), standard errors (SE), R^2, adjusted R^2 and p values from the hierarchical regression predicting distance to optimal scores for narrow and wide priors using vocabulary knowledge, verbal (Backwards digit span; BD) and visuospatial working memory (Corsi memory span).

		Variable	β	SE	p	R^2	Adjusted R^2
Narrow prior	Step 1	Vocabulary	.002	.005	.720	.002	-.014
	Step 2	Vocabulary	.001	.006	.862	.006	-.027
		BD	-.009	.017	.611		
	Step 3	Vocabulary	.001	.006	.917	.032	-.017
		BD	-.011	.017	.516		
		Corsi	.040	.032	.217		
Wide prior	Step 1	Vocabulary	.001	.003	.534	.006	-.010
	Step 2	Vocabulary	.002	.003	.462	.010	-.023
		BD	.004	.009	.636		
	Step 3	Vocabulary	.001	.003	.584	.016	-.034
		BD	.003	.009	.691		
		Corsi	.010	.017	.544		

None of the predictors were correlated with distance-to-optimal scores.

Having established that none of the predictors uniquely contributed to explaining the variability in distance-to-optimal scores, we next sought to determine whether each of these predictors is at all correlated with how close to optimal an observer is across the two priors. We computed the Pearson correlation between each predictor separately for the narrow (Figure 4.4) and wide prior (Figure 4.5). Distance-to-optimal scores were not significantly associated with visuospatial working memory capacity when the prior was narrow or wide (see Figures 4.4A and 4.5A, respectively). Also, no significant relationship was found for either of the prior variances between distance-to-optimal and verbal working memory capacity (Figures

4.4B and 4.5B) and distance-to-optimal and vocabulary knowledge (see Figure 4.4C and 5C).

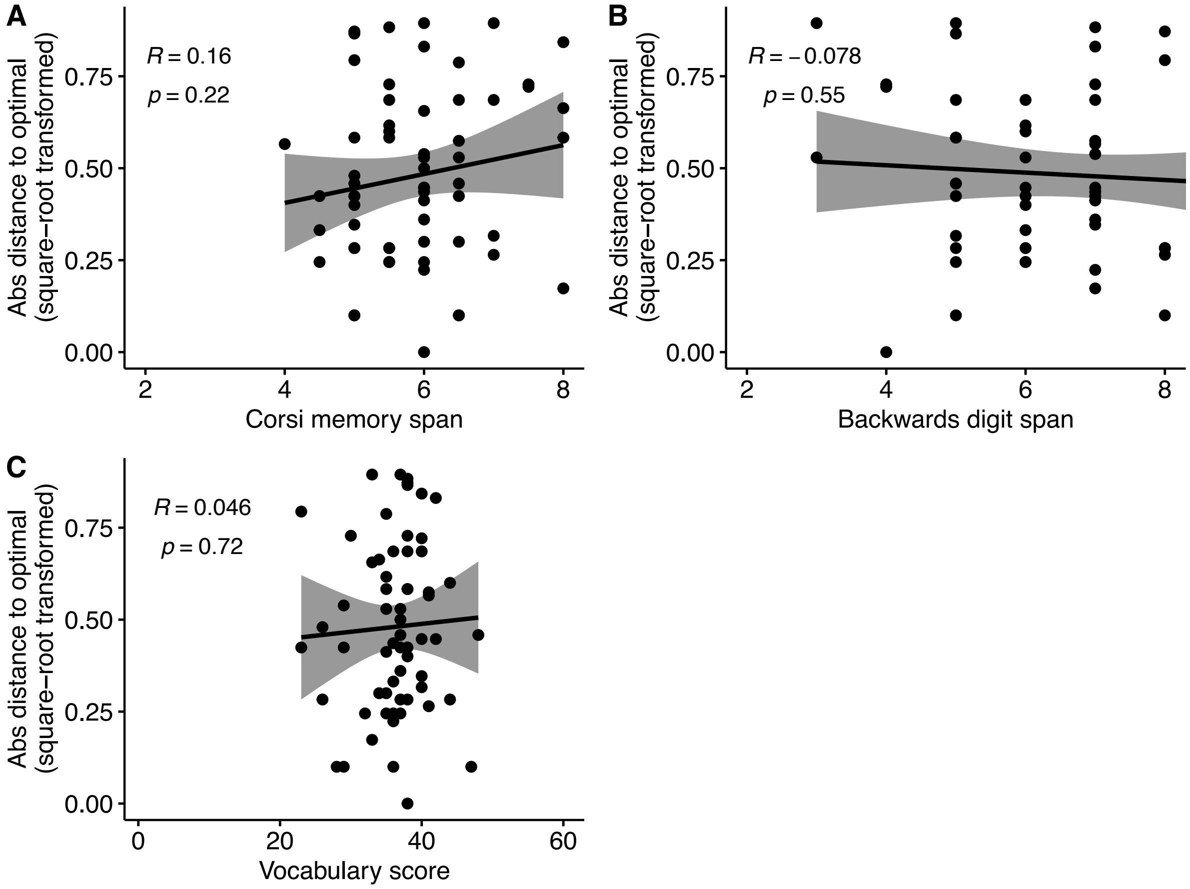

Figure 4.4. Correlations between distance-to-optimal scores (narrow prior) and visuospatial working memory as measured by the Corsi blocks test **(A)**, verbal working memory as measured by the Backwards-digit span **(B)** and vocabulary score **(C)**.

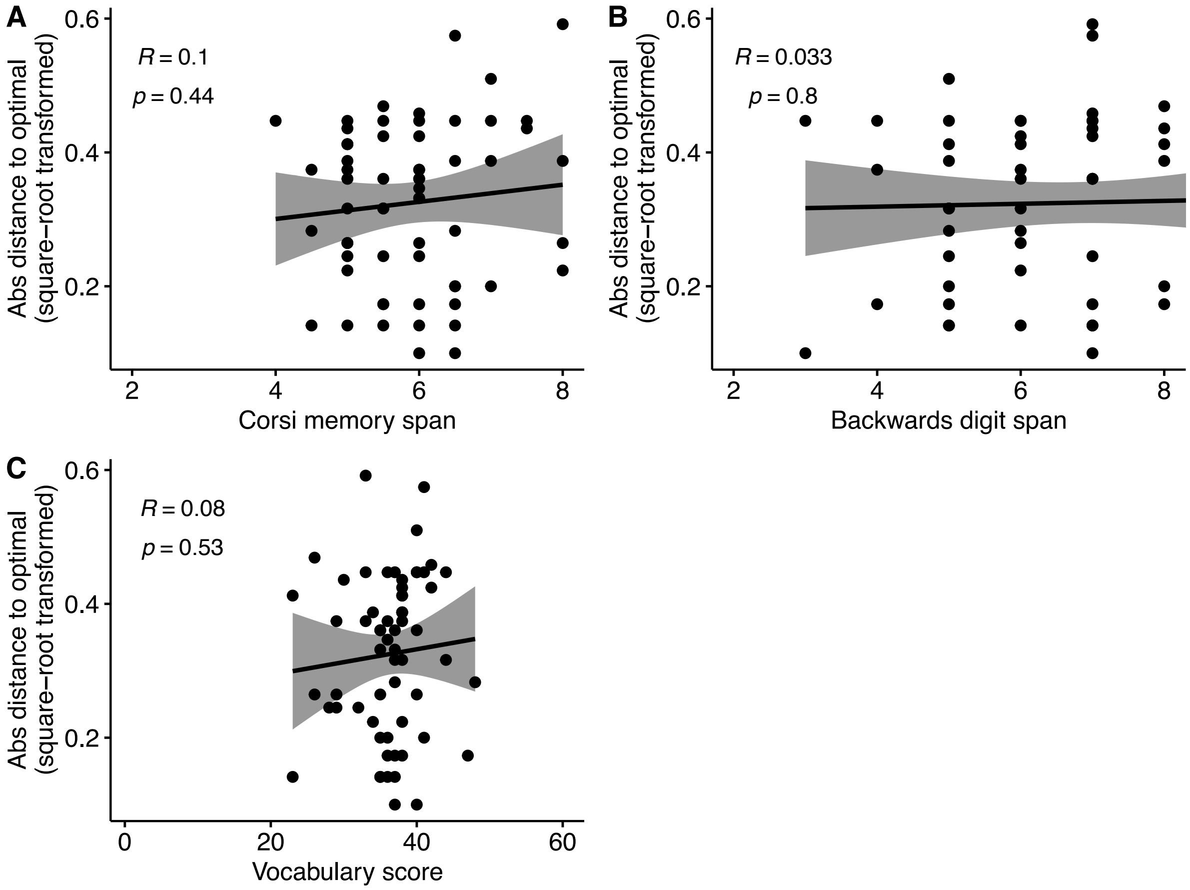

Figure 4.5. Correlations between distance-to-optimal scores (wide prior) and visuospatial working memory as measured by the Corsi blocks test **(A)**, verbal working memory as measured by the Backwards-digit span **(B)** and vocabulary score **(C)**.

4.4. Discussion

In this study we set out to ascertain whether visuospatial working memory explained unique variance, beyond that explained by verbal working memory and vocabulary knowledge (as a proxy for general IQ) in how close to optimal observers were in a localisation task we used before (Chapter 3). We replicated our previous finding of adults changing their reliance on the prior 1) dependent on its reliability, and 2) over time (Experiment 1, Chapter 3). However, individual differences in any of the measures, visuospatial working memory, verbal working memory or vocabulary knowledge - did not explain significant variance in distance-to-optimal indices for either of the priors.

As mentioned above, we duplicated our previous finding of adults 1) weighted the prior according to its relative reliability, assigning less weight to the less reliable prior overall, and 2) increased their reliance on the prior as they experienced more trials. We replicated this effect in a sample size more than twice that of typical psychophysics experiments (roughly $N = 30$, some with as low as seven participants, Berniker et al., 2010). These data support the conclusion that adults take reliability into account (as ideal observers would) but generally, do not do so optimally.

Combining psychophysics together with cognitive measures allowed us to test whether individual differences in working memory and general IQ (with vocabulary knowledge used as a proxy) explain variance in how close to optimal individual observers are. However, visuospatial working memory capacity, as measured by the Corsi memory span, did not uniquely contribute, beyond verbal working memory or general cognitive ability. A series of studies demonstrate that parietal regions take part in decision making, especially for decisions that require accumulation of evidence (Morcos & Harvey, 2016; Schott et al., 2019). Since the parietal cortex has also been shown to contribute to the storage of working memory information (Romo et al., 1999; Shadlen & Newsome, 2001), we expected a relationship between working memory and perceptual decision-making. In fact, a positive correlation of working memory and incidental learning of probabilistic sequences (Cashdollar et al., 2017; Park et al., 2020) and the rate of evidence accumulation (Ester et al., 2014; Schmiedek et al., 2007) has also been reported. So why did we not find a relationship between working memory and efficient decision-making, when so many links have been reported previously?

The storage capacity of working memory should play a central role in allowing information encountered further back in the past to be maintained in memory. In doing so, memory capacity may be critical for building the overall prior distribution (i.e., *learning* the prior), but should not be needed to learn how to *use* the prior efficiently. This raises the possibility that adult observers had learned the prior distribution equally well (which would explain why there was no effect of working memory), but what they are variable in is how good they are at assigning appropriate weights to the prior, relative to the sensory information (which would explain why we still observed considerable variation in how close to optimal observers were).

Distance from optimal as a measure of efficient use of novel statistics

Another possibility is our choice of outcome variable. We used distance from optimal as a proxy for efficient use of novel task statistics. We hypothesised that higher working memory capacity, particularly visuospatial, would increase the amount of past information that could be stored or processed efficiently, steadily shortening the distance from optimal predictions. Distance from optimal thus allowed us to determine how *efficiently* observers use the novel statistics.

Nevertheless, there is the possibility that there was actually a relationship between distance from optimal and working memory, but our outcome variable might not have been sensitive enough to detect such a relationship. In fact, previous studies that did report a positive correlation have used more sensitive measures, such as reaction times or computational modelling (Ester et al., 2014; Schmiedek et al., 2007). Using computational modelling, for example, Ester et al. (2014) was able to determine that working memory is linked to drift rates, i.e., the rate at which evidence is accumulated over time, but not response threshold (i.e., amount of sensory evidence needed to reach a decision). It is therefore possible that our outcome measure was useful but provided relatively little information about what aspects of the decision-making working memory plays a role in. Future studies could use more sensitive measures to clarify the contribution of working memory to different parts of the decision-making process.

Corsi blocks as a measure of visuospatial working memory

Lastly, it is possible that that the Corsi blocks task was not a sensitive enough measure to detect individual differences in distance-from-optimal scores. Corsi blocks involves remembering a number of discrete locations in a sequence. In contrast, in the localisation task, the targets were at different locations, drawn from a continuous spatial distribution. Thus, it is possible that because the Corsi-blocks task involves a more discrete form of spatial memory (remembering up to nine distinct locations in a sequence), it was not sensitive enough to detect inter-individual variations where the distribution of locations was continuous across space. Future studies will need to use measures which provide a more sensitive means for capturing individual differences in a continuous localisation task like ours, such as delayed estimation (Wilken & Ma, 2004) which also uses a continuous response scale.

Does our task require verbal working memory?

Verbal working memory also had no impact of how close adult observers were to optimal, which is what we would expect if verbal working memory was not required for the localisation task. It appeared possible that observers may have internally verbalised the spatial order of the targets (e.g., where the next target would likely appear, relative to the one they have just seen) - a process which would, indeed, rely on verbal working memory. However, it is unlikely that such verbal strategies were used over the *entire* experiment. This may explain the lack of a relationship between verbal working memory and distance-from-optimal indices in our study. Nonetheless, further studies would be required to completely rule out the possibility that verbal working memory is involved in learning the task statistics.

The role of general cognitive ability

Although we observed considerable variability in our general cognitive measure (vocabulary knowledge; WASI-II), we did not find that this variability explained differences in how close to optimal different individuals were. This result is not surprising given that all of our participants were university students, and as a consequence, were potentially quite homogenous in terms of education, socioeconomic status and general cognitive capacities. We acknowledge that based on our results, we cannot confirm whether this result would hold for the general population, but this could be addressed in future work.

Violation of key assumptions

We checked the data for violations of key regression assumptions, such as residual normality and homoscedacity (homogeneity of variance). The residuals were not normalised: this is why a square root transformation was applied for the distance-from-optimal scores (Bishara & Hittner, 2012). An alternative would be to use a non-parametric test (Miller, 1986) or generalised linear models where the distribution of error can be changed from normal to other error distributions, e.g., Poisson or binomial (O'Hara, 2009). However, we believe that the latter, in particular, is more problematic than transforming the data because certain models (e.g., Poisson) are very sensitive to violations of their normality assumptions, increasing the risk of Type I error (Ives, 2015; Warton et al., 2016) (as was argued in Knief and Forstmeier, 2021).

To summarise our results, we did not find that visuospatial working memory predicted individual differences in how close to optimal observers were, beyond what is already explained by verbal working memory and general cognitive abilities, and regardless of whether the prior was narrow or wide. Future work using computational modelling and more sensitive measures will confirm with a greater deal of certainty if and how working memory contributes to perceptual decision-making.

Chapter 5

General Discussion

5.1 Introduction

In this thesis, we sought to investigate how observers learn and use novel priors efficiently, what the limits of these abilities are, how they change with age and what factors that may underlie these abilities. Across three experiments in Chapter 2, we examined whether adults' ability to generalise learned priors to untrained cue reliabilities is limited by task complexity. Across three experiments in Chapter 3, we then looked at how adults and children adapted to novel statistics and probed some of the factors that may affect the extent and rate of this adaptation. These experiments were supplemented with an individual differences approach, examining whether individual variation in working memory capacity relate to how close to optimal adults are (Chapter 4).

Firstly, we will summarise the central findings of this thesis (Section 5.2). We will then explain how these findings build and expand on findings in the existing literature (Section 5.3). Then, we will outline the potential implications of our findings (Section 5.4), their limitations and possible avenues for future research (Section 5.5).

5.2 Summary of main findings

5.2.1 Chapter 2

The existing literature does not currently tell us whether observers can immediately transfer probabilistic information in situations that more closely resemble the complex and dynamic environments in which humans operate – i.e., where statistics can change suddenly. Therefore, across three experiments in Chapter 2, we used a task which was made more complex than previously used (Sato & Kording, 2014) by interleaving multiple prior and sensory variances. We asked whether adults would *immediately* generalise already learned priors to a new level of sensory uncertainty (demonstrated by instant change in cue weighting; "Bayesian transfer" (Maloney & Mamassian, 2009)), which would be consistent with Bayesian inference, or whether they would take longer to adapt to the untrained sensory reliability (suggestive of a mechanism similar to reinforcement learning). In Experiment 1, we found only weak evidence of transfer, with no differences in weights when sensory uncertainty jumped from medium to high. In Experiment 2, where participants were explicitly told of the different prior variances, we found evidence of transfer, with significantly less weight assigned to the high, relative to the medium, sensory uncertainty. In both Experiment 1 and Experiment 2, performance was suboptimal. This was also the case in Experiment 3, where all prior and sensory variances were included from the beginning (no transfer). We also found that a model that weights sensory information by taking into account only its internal noise provided a better account of the data than a model that accounted for the external noise only and a model that accounted for both internal and external noise in the sensory cue. This is the first study (so far as we are aware) to examine abilities to transfer in dynamic and complex environments.

5.2.2 Chapter 3

Despite overwhelming evidence that humans detect and learn novel statistical regularities as early as a few months after birth, recent studies have shown that children do not begin to use these efficiently until much later (around 9-10 years of age; Bejjanki et al., 2019; Chambers et al., 2018). However, the factors underlying the long development of this critical perceptual skill have not been explored. Across three experiments in Chapter 3, we probed several factors which likely affect the ability to efficiently use the statistics. In Experiment 1, we traced the detailed time course of how children and adults adapted to novel changing task statistics. We found that 6- to 8-year-olds learned enough about the two priors to weight them differently, but this learning was so slow that, unlike in 9–11-year-olds and adults, it did not show measurable changes over time. In Experiment 2, where all previous trial outcomes were explicitly available, 6- to 8-year-olds learned to use the statistics faster and more efficiently (i.e., became adult-like, moving closer to optimal), and adults' performance became indistinguishable from optimal values. Experiments 1 and 2 are the first to directly compare adults and children's weightings when the prior was directly observable and, therefore, did not need to be learned (Experiment 2), as opposed to having to be inferred from past events (Experiment 1). Additionally, our results from Experiment 3 have shown for the first time that when several weighting ‘rules’ had to be stored and managed at the same time, 6- to 8-year-old children’s abilities to distinguish between statistical distributions tended to break down, and adults’ judgments became more child-like. Overall, these findings show that children are sensitive to novel task statistics from a young age but making efficient use of these for perceptual decision-making under uncertainty depends on development of central cognitive abilities, including working memory and cognitive flexibility. Our results suggest how early sensitivity to task statistics may be reconciled with much later development of abilities to use these statistics efficiently during perceptual decision-making.

5.2.3 Chapter 4

Having observed that adults’ performance became indistinguishable from optimal when working memory demands were relaxed (Experiment 2, Chapter 3) but not otherwise (Chapter 2; Experiment 1, Chapter 3), we intended to deepen our understanding of this result. In Chapter 4, we took a different approach by examining whether inter-individual differences in working memory capacity are related to how close adults are to optimal. We focused on visuospatial working memory due to the spatial nature of our task. We also included a measure of verbal working memory and vocabulary knowledge as control measures. Using a sample size more than three times higher than that used in our previous experiments (63 adults vs. 19 in each experiment in Chapter 3), we replicated our findings of adult observers changing their reliance on prior, according to its reliability, and over time (Experiment 1, Chapter 3). However, individual differences in visuospatial working memory, verbal working memory or vocabulary did not explain significant variance in distance-from-optimal indices for either of the prior variances. As far as we are aware, no studies to date have examined the relationship between distance-to-optimal indices and working memory capacity.

5.3 Contributions to the literature

5.3.1 Limited by complexity

A strong finding emerging from Chapters 2 and 3 is that the ability to rapidly and flexibly perform Bayesian inference is limited in more complex, changing environments. Abilities to generalise to untrained reliabilities is a powerful test of Bayesian inference ("Bayesian transfer"; Maloney & Mamassian, 2009). Using this approach in a more complex task than has been used previously (e.g., Sato & Kording, 2014), in Experiment 1 (Chapter 2), we were able to show that adult observers did not generalise learned priors of stimulus locations to a new cue reliability level (Maloney & Mamassian, 2009). This contrasts with Sato and Kording (2014), who used a far simpler visual estimation task and showed that adult observers immediately transferred what they had learned about likelihood variances to a new prior. The fact that in our study, observers did not generalise and remained far from optimal - despite receiving feedback - suggests that complexity poses a fundamental constraint on performing Bayesian inference. This conclusion makes sense because we know that performing accurate inference is computationally costly (e.g., time, memory); therefore, more complex tasks that draw to a much larger extent from memory and other resources may render performing optimal inference difficult or unfeasible (Beck et al., 2012; Ma, 2012).

Importantly, in this thesis, we also show that complexity similarly poses a constraint on children's performance. We showed that 6- to 8-year-old children can differentiate between the prior variances, but only when there was a single level of sensory uncertainty, making the task much simpler and easier (Experiment 1, Chapter 3). However, when this task was made more complex by interleaving two levels of sensory uncertainty (in addition to changing prior uncertainty halfway through the task), children of the same age did not differentiate between the narrow and the wide prior (Experiment 3, Chapter 3). Our results corroborate earlier work by Bejjanki et al. (2019), who showed that 6-7-year-old children changed their weightings when only the sensory uncertainty changed, but not when both sensory and prior uncertainty changed in an interleaved fashion (multiple 'rules' at the same time, making the task more complex). Similarly, Chambers et al. (2018) found that children aged between 6 and 8 years were less sensitive to changes in prior uncertainty than older children and adults when both prior and sensory uncertainty changed.

Together, our findings speak directly to a long-standing question in the field of what the limits of Bayesian inference are. Quite consistent with the role of complexity in limiting optimal inference, both adults and children's performance became worse in more complex tasks.

What does this tell us about real-life decisions?

Exploring whether people make the best, most efficient use of newly learned information in more natural and complex tasks is crucial to understand how people handle real-life situations, where the reliability of sensory inputs changes at unexpectable times. In simple, stable environments, you could find the optimal solution by learning simple rules, such as reinforcement learning. However, these would not generalise if the environment changes. By contrast, efficient, Bayesian inference allows for flexible behaviour in changing environments. Studies using more natural and complex tasks and stimuli therefore provide a crucial test of what people learn when they learn to make efficient perceptual inferences. Given the ubiquity of the non-stationarity of our environments, our results highlight the need for developing interventions that are aimed at helping observers to better detect when the environment changes, and to adjust their decisions accordingly.

5.3.2 Adult-like integration of evidence from various sources after 8 years, but earlier if you know more about the underlying structure

Although some basic abilities, such as extracting statistical regularities from sensory input, exist in infants (Fiser & Aslin, 2002; Kirkham et al., 2002; Saffran et al., 1996), the ability to make efficient use of novel statistical regularities does not become adult-like until late childhood. We showed this in Chapter 3, Experiment 1; 6- to 8-year-olds changed their reliance (weighting) on the prior when its reliability changed but the extent and the rate of their adaptation to the task statistics was far from adult-like. Chambers et al. (2018) also found a lack of adult-like weighting in the same age group. By tracing the time course of adaptation in adults and children between the ages of 6 and 11 years, we can now add that although the youngest children learned enough about the priors to treat them differently, this learning was so slow that, unlike in 9–11-year-olds and adults, it did not show measurable changes over time. And indeed, previous reports of lack of reliability-dependent changes in prior weighting in young children (Chambers et al., 2018) could result from such slow learning. This slowness contradicts the idea that young children *cannot* use novel statistics in an efficient, adult-like way, and instead aligns more closely with the idea that children are extremely slow (much more than adults) to learn to make efficient use of these statistics. This conclusion is supported by studies from Manning et al. (2020) and Ratcliff et al. (2012) who used computational modelling to show that young children accumulate evidence at a slower rate than adults.

Our results also show that the ability to make more efficient, adult-like use of novel statistics can emerge already by the age of 6, provided the underlying task structures are made explicit. We show this in Experiment 2 (Chapter 3), where we made all previous trial outcomes (aka the prior) explicitly available and found that 6- to 8-year-olds' adapted to the statistics in similar ways to adults, when they had not before (Experiment 1, Chapter 3). The difference in performance of 6- to 8-year-olds between Experiments 1 and 2 indicate that the 6- to -8-year-olds in Experiment 1 may not have adapted to novel statistics in a similar way to adults because they were uncertain about what those statistics are. We see a parallel to the literature on the development of cue combination. In Section 1.3.6.3, we discussed findings of children younger than 8 years of age not combining multiple sensory cues in an efficient, adult-like way, possibly because they are unsure about whether the cues are from the same or different objects (Dekker & Lisi, 2020; Rohlf et al., 2020). This idea – that young children may be unsure about the underlying causal structure - also gains support from a recent study by Negen et al. (2019); seven-, eight-, nine-, and ten-year-old children were presented with an auditory and a visual cue, and told that both of these cues indicated the location of an unseen target. The cues were from the same location and the correct response was shown at the end of each trial. Rohlf et al. (2020), on the other hand, asked participants to localise sounds without feedback; also, there was a spatial separation between the visual and auditory signals. So, while there is no doubt of what the causal structure was in Negen et al.

(2019) (visual and auditory cues originate from the same source), the causal structure in Rohlf et al. (2020) was not certain.

Coming back to our own developmental work, we reiterate that in Experiment 1 (Chapter 3), the underlying task structure was unknown, and had to be learned from feedback; 6-8-year-olds adapted extremely slowly. In Experiment 2 (Chapter 3), we showed the participants how the targets were distributed, and thus, reducing the ambiguity in the underlying structure of the task and found adult-like adaptation in children as young as 6 years. Taken together, these findings provide strong evidence that the real bottleneck to children combining optimally is not so much in them being incapable of doing so, potentially because of immature neural mechanisms, but in uncertainty about the statistical structure of the environment.

We want to point out that knowing more about the statistics and structure of the task also improved adults' performance. Presenting the prior distribution explicitly brought adults' performance closer to the optimal values in Experiment 2 (compared to Experiment 1, Chapter 3), such that it was now indistinguishable from optimal. Also, in Chapter 2 we found that adult observers (1) moved closer to optimal, and (2) generalised to the untrained likelihood variance when we explicitly told them that there were two prior variances, one more reliable than the other (Experiment 2); no evidence of transfer was found when this information was not communicated to participants (Experiment 1). Similarly, Acerbi et al. (2014) reported that adults could compute with complex priors that change on every trial when they were explicit but not when the prior distribution(s) had to be learned (Acerbi et al., 2012; Körding & Wolpert, 2004). These results reinforce the notion that knowledge of the underlying statistics is essential in performing Bayesian inference, to the point that without this knowledge, adults struggle to generalise across contexts, and children take longer to learn to use information in an efficient, statistically optimal way.

5.3.3 Limited resources

As well as incomplete knowledge about the statistical structure, our findings of (1) optimal use of novel statistics in adults, and (2) greater, much more adult-like adaptation to these statistics in young children (Experiment 2, Chapter 3), can be explained equally well by resource-rational theories (Bejjanki & Aslin, 2020; Lieder &

Griffiths, 2019). According to this account, there is a fixed amount of resources (e.g., storage capacity of working memory) that could be used to learn and efficiently adapt to novel statistical priors, especially in children, when working memory and other executive functions are still developing (Best & Miller, 2010). In showing all previous trial outcomes on-screen, we made the prior distribution known while at the same time relaxing demands on working memory, freeing up cognitive resources. This may be why the 6- to 8-year-olds in Experiment 2 were adult-like in how they adapted, whereas the children in Experiment 1 were not; it is possible that their working memory capacity did not allow them to retain accurately the preceding target locations needed to correctly infer the underlying structure of the task (Chapter 3). These findings are also consistent with parallel improvements in working memory during childhood (Zelazo et al., 2008). Interestingly, adults' performance also became indistinguishable from optimal (Experiment 2, Chapter 3). That being said, we know that working memory is a limited resource even in adults (Bays & Husain, 2008) regardless of whether observers 1) remember each individual location, or 2) instead maintain a summary representation of what the distribution looks like and update that with new samples to more efficiently do the task. These processes can be difficult to distinguish (Hansmann-Roth et al., 2021; Spicer et al., 2020; Sun et al., 2019), but both require accurate representation and updating in working memory.

Our results from Experiment 3 are also consistent with the resource-rational explanation: when multiple interleaved levels of sensory uncertainty were introduced, increasing demands upon cognitive resources, the youngest children's abilities to distinguish between statistical distributions broke down, and adults' judgments also became more child-like (Chapter 3). This result is consistent with Chambers et al. (2018), who also found this lack of reliability-weighting of the prior in a situation with multiple sensory likelihoods. Also, Bejjanki et al. (2019) showed that 6-7-year-old children changed their weightings when only the sensory uncertainty changed (i.e., single 'rule), but not when both sensory and prior uncertainty changed in an interleaved fashion (multiple 'rules' at the same time). Also in line with our findings is evidence that cognitive flexibility improves at older ages. Irwin-Chase and Burns (2000) found that the ability to manage and switch between different rules improves between the ages of 5 and 8 and becomes adult-like after 8 years of age. Other studies have also shown that young children find multiple rules extremely difficult to manage (Anderson et al., 2011; Harada et al., 2018; Huizinga & Van Der Molen, 2007; Zelazo et al., 1996, 2008).

Altogether, our developmental results support the notion of a resource-rational strategy where people economise on the limited cognitive resources they have available to learn about the statistics of the environment. Our conclusion - that understanding decision making in humans, particularly as they grow older, requires that we account for limitations – is in keeping with a long-standing tradition of understanding development in terms of cognitive limitations, such as attention and working memory (Klahr, 1992; Klahr & Macwhinney, 1996).

We note that the two different explanations – incomplete knowledge of the underlying statistics and resource rationality – are not mutually exclusive. In fact, finite resource could explain why young children may be more uncertain about the same environments that adults are certain about. If children's abilities to learn about and represent the novel statistics are more limited – presumably because they have less resource available (e.g., lower memory capacity) – they may be still learning what the underlying statistics of the environment are.

5.3.4 Neural explanations

Although our data are compatible with many perspectives, the possibility we may potentially be able to rule out is that there is an in-principle inability of younger children to combine the relevant information in a way that is weighted by reliability. Such an inability might be expected if there were a central immaturity in the brain's implementation of weighted averaging. For example, it has been suggested that there is a reduction in the amount of inhibition that occurs through divisive normalisation and as a consequence, neurons have a smaller impact on the activity of nearby singe neurons (Carandini & Heeger, 2013): if true, this would result in underweighting of priors, which has been well documented in adults with autism. If we assume, in line with this explanation, that children younger than 9 years of age have reduced divisive normalisation, children in this group should systematically underuse priors, compared to adults. After this age, children should be able to learn and make efficient use of novel statistics in a similar way to adults. While the experiments reported in this thesis were not designed to test this theory, the experiments in Chapter 3 seem to contradict it, with adult-like adaptation to novel task statistics at the age of 6 years as the demands on memory capacity are reduced (Experiment 2, Chapter 3). Similarly, when there were multiple 'weighting' rules, 6-8-year-olds became less sensitive to changes in the reliability of the prior. Therefore, reduced divisive normalisation does not do a good job of accounting for the less weighting of the prior in younger age groups.

In sum, we believe we have made several novel contributions to the literature on the development of perceptual priors. Firstly, similarly to previous studies (Bejjanki et al., 2019; Chambers et al., 2018) we did not find evidence for adult-like use of novel statistics before 9 years of age in Experiment 1 (Chapter 3). These results are also consistent with what has been found in the vast majority of cue combination studies in children (Gori et al., 2008; Nardini et al., 2008, 2010): children younger than 8 years do not combine multiple sensory cues in a way that gives less weight to the less reliable cue. However, more recently, we and others (cue combination; Rohlf et al., 2020) have shown that the ability to combine across cues and with prior knowledge emerges in young children from the age of 5, provided the environmental structure is known, thereby freeing up resources and capacity for learning how to combine information optimally. This is a shift away from theories, suggesting that the basic mechanism needed to combine information optimally does not develop until 9 years of age (Chambers et al., 2018).

Secondly, the fact that when we made the prior explicit, and thus removing the need to learn it, the performance of 6- to 8-year-old children reached adult-levels and moved closer to optimal, provides compelling evidence that previous reports of young children's failures to weight the prior efficiently and in an adult-like way (Chambers et al., 2018) are more likely a result of difficulties with *learning* the prior, as opposed to *learning how to use* it. Although this seems to go against the long-held belief that children learn novel statistics from a young age (Fiser & Aslin, 2002; Kirkham et al., 2002; Saffran et al., 1996), it is worth noting that the infancy studies held as evidence that infants are already good at learning new statistical distributions only show whether children can discriminate between familiar and novel stimuli above chance; these studies do not tell us how *accurately* these distributions are learned, compared to older children and adults. Incidentally, accuracy of learning would be hard to assess with those incidental looking-time methods, as there is no explicit task pushing people to be as accurate as possible.

5.4 Implications

5.4.1 Shared underpinnings with cue combination

While cue combination and prior and sensory combination are both key aspects of Bayesian inference, it is not known whether they are linked and are underpinned by a common mechanism. Our findings mark an important step in linking the development of cue combination and perceptual priors. Evidence for efficient, adult-like adaptation to novel prior statistics (Experiment 1, Chapter 3; Bejjanki et al., 2019; Chambers et al., 2018) seems to emerge around the same age when children start to combine multiple sensory cues efficiently (Gori et al., 2008; Nardini et al., 2008, 2010). This observation has led many to believe that there is a common neural mechanism underlying both cue combination and integration of priors, perhaps to do with representing and / or computing with probabilities (Ma, 2012), or the neural mechanism needed to integrate cues (and prior knowledge) in a reliability-weighted way (Ohshiro et al., 2011). Our own studies and others (Rohlf et al., 2020) now move away from such neural explanations, instead suggesting a common problem with uncertainty about the underlying structures, or a much more basic, more priors-specific problem of coping with excess demand on immature executive functions (Best & Miller, 2010; Diamond, 2013; Gur et al., 2012). More studies are now needed to determine to tease apart these alternatives.

5.5 Future Directions and Limitations

5.5.1 The importance of testing beyond 11 years

Studies that extend beyond 11 years into adolescence would help determine the specific age at which integration of prior knowledge becomes adult-like. In

Experiment 1, Chapter 3, we found less weighting of priors in children, compared to adults, with no significant differences between the 6-8-year-olds and 9-11-year-olds. This means that learning and/ or the efficient use of novel priors is not adult-like even by the age of 11 years, suggesting that these abilities continue to mature into adolescence. However, to our knowledge, research on integrating and weighting information by its reliability in adolescence is limited (Haller et al., 2018). Adolescence is also a time when working memory (Experiment 2, Chapter 3) and other executive functions that efficient prior integration may depend on, e.g., storing and switching between multiple ‘rules’ (Experiment 3, Chapter 3) also improve (Best & Miller, 2010; Gur et al., 2012). Clearly, given the ties between the delays in efficient use of novel statistics and the late maturation of executive functions and the brain regions they rely on (Baum et al., 2017; Gu et al., 2015; Hagmann et al., 2010; Huang et al., 2015), future studies should examine longitudinal changes over time between early childhood, through early teens and into adulthood.

5.5.2 The importance of using computational models

Across the three experiments in Chapter 3, we demonstrated differences in the learning and efficient use of novel statistics between adults and children aged 6 to 11 years. However, we did not investigate the mechanisms underlying these abilities, or the decision strategies adults vs. children use when performing perceptual decision-making tasks. Computational modelling lends itself well to these challenges. In the cue combination literature, this approach – of fitting models to experimental data – has resulted in a number of insights into the mechanisms and algorithms underlying behaviour (Nardini et al., 2008; Rohlf et al., 2020). Using this approach, Nardini et al. (2008) showed that adults’ behaviour was most consistent with an integration process that computes a reliability-weighted average; by contrast, the behaviour of 4- to 5-year-olds and 7- to 8-year-olds was better explained by a process that switches between the cues. More recently, model comparison demonstrated that a model which accounted for ‘causal uncertainty’, i.e., the uncertainty about whether the cues come from one object or different objects fits both adult and children’s data better than a model that always integrates or a model that always kept the cues separate (Rohlf et al., 2020). Also, by varying the parameters of the model, Rohlf et al. (2020) demonstrated that young children are less likely to assume a common source. In sum, by using computational modelling, Rohlf et al. (2020) was able to demonstrate that young children can combine cues but are generally uncertain about whether the cues come from the same source. And indeed, we could also use computational modelling to better understand why young children in our studies tend to underweight the prior. For example, it has been suggested that adults with autism, who similarly to typical young children also underuse priors, do so because they overestimate the volatility in a changing environment (Lawson et al., 2017). Behrens et al. (2007) makes sense of this observation by saying that when the environment is believed to change very often, information from the past (i.e., priors) is not the most informative, and will therefore be weighted less. It is therefore possible that much like the lower prior expectations of a common source shown by Rohlf et al. (2020), young children could also have lower prior expectations that the environment will remain stable. By looking at how different model parameters, such as the *a priori* expectation of environmental stability, vary with age, we could better understand the reasons behind the under-use of priors in young children. All of that being said, distinguishing reliably between models and parameters can be very challenging with developmental data (Scarfe & Glennerster, 2018) that is generally more prone to response biases and noise (Jones & Dekker, 2017; Manning et al., 2018).

Computational modelling approaches have also been applied to model the *learning process* itself (Wilson & Collins, 2019). This is because each trial depends on the preceding trial(s); therefore, the ‘classic’ analysis of averaging across trials and conditions may not be as informative. In fact, we could apply computational models to our behavioural data to find out whether young children adapt more slowly to the prior (1) but will eventually reach the same asymptote as adults, or (2) would plateau before reaching the adult / optimal asymptote. The number of trials in the experiments in Chapter 3 was not sufficient to determine whether 6-8-year-olds, 9-11-year-olds and adults would all converge to a common asymptote with enough exposure; however, in the future, probabilistic learning tasks could be modelled with a simple exponential rule with parameters for the learning rate, asymptote and the initial weight given to the prior. Observing how the parameters of the specific components of the model (e.g., learning rate) vary with age would help us to better understand which elements of the model are still developing in children.

We could also use computational modelling to quantify how working memory contributes to learning. Although simple behavioural analyses could explain the broad qualitative patterns we saw in Experiment 2, Chapter 3, it could not tell us which mechanisms are more likely to underlie these patterns. Limitations in working memory capacity could result in remembering previous locations imprecisely or forgetting locations with time (i.e., recency effects). Models that make precise, quantifiable predictions would help us better understand and quantify the effects of working memory limitations, particularly in the context of learning and decision making.

5.5.3 Neural mechanisms that might underlie attenuated priors and the importance of eye-tracking methods

In Chapter 3, we reported less weighting of the prior in relation to sensory input in 6- to 8-year-olds. However, we did not explore the neural mechanism(s) that may be underlying the underuse of priors in this age group. Reduced inhibition that occurs through divisive normalisation is believed to lead to priors being underweighted in adults with autism (Rosenberg et al., 2015). Divisive normalisation has further been implicated in the modulatory effects of attention and multisensory integration on neural responses (Ohshiro et al., 2011; Reynolds & Heeger, 2009). These findings are consistent with the idea that there may be a link between neural mechanisms, specifically involving divisive normalisation, and underweighting of priors in young children. This should be a major consideration for future studies of the development of novel priors.

As well as MRI measures which could be used to uncover underlying neural mechanisms like reduced divisive normalisation, we could also use eye-tracking methods to better understand and interpret our results. Studies have employed eye-tracking to take pupil diameter measures, shown to encode the trial-by-trial level of surprise at the time that feedback was provided (Lavín et al., 2014; Preuschoff et al., 2011). These pupil measures can, for example, help elucidate whether young children in Experiment 1 (Chapter 3) took longer to change their reliance on the prior when it changed in reliability because they did not notice relevant feedback, or because they did not learn from it. Eye-tracking could also be used to probe what is it that people had actually learned about novel statistics. Behavioural measures alone cannot tease apart suboptimality due to not representing the correct distribution, or not combining it correctly with current sensory input. One way to do this is to occasionally show stimuli that are outliers (e.g., show a target in a location that falls outside the range of the distribution, so it is unexpected). If people struggle with *representing* the distribution, we might expect to see increases in pupil size from surprise signals generated by such outlying observations. Future studies should attempt to repeat our findings but using eye-tracking methods.

5.5.4 Individual differences

In Chapter 4, we found that distance-to-optimal indices in adults cannot be predicted by individual differences in visuospatial working memory capacity; and neither by verbal working memory capacity or vocabulary knowledge. This is despite evidence from work by us (Experiment 2, Chapter 3) and others (Bagneux et al., 2013; Cui et al., 2015; Duncan et al., 2012; Pereg & Meiran, 2019) that working memory is crucial for learning new task rules and making decisions under uncertainty. Broadly, we concluded that is it possible that unlike in development, in adults, there are no longer important differences in working memory capacity that limit performance on our localisation task. This suggests that adults have learned the novel statistics similarly well, but are variable in other factors, e.g., how good they are at assigning appropriate weights to the prior and sensory information. Alternatively, it is possible that in order to capture individual differences in working memory capacity, we need to use a more sensitive measure. As our measure of visuospatial working memory, we used the Corsi blocks task (Corsi, 1972), which is one of the most widely used paradigms in visuospatial working memory research. In this task, the participant must hold a sequence of up to nine discrete locations in memory and then recall the sequence of locations in the correct order. By contrast, our task involves remembering locations which are continuously distributed. Could it be that the Corsi blocks test was not sensitive enough to detect inter-individual variations where the distribution of locations was continuous across space? An example of a sensitive measure of visual working memory is the delayed estimation task (Ma et al., 2014), which uses a continuous response space, rather than discrete ‘locations’. Future studies should establish whether using more sensitive measures, like the delayed estimation task, can detect differences in working memory capacity and predict individual differences in how close to optimal an observer is.

Another consideration in individual differences research is the homogeneity of the sample. Our participants were similar in education, socioeconomic position, and very likely, in general cognitive ability. Therefore, in the future, studies should explore individual differences in the general population.

5.6 Conclusions

The aim of this thesis was to gain a better understanding of how people learn and make efficient, statistically optimal use of novel statistical distributions. We have shown that increasing environmental complexity complicates performing Bayesian inference. We have also shown that making the environmental structure more explicit has a key role in our ability to use novel statistics efficiently in both children and adults, with children's performance reaching adult-like levels and adults' performance matching that of optimal decision makers. Lastly, potentially ruling out individual variation in working memory capacity as a source of variation of adults' distance from optimal opens up avenues to 1) replicate our finding but with more sensitive measures, and 2) investigate other factors which may impact how close to optimal adults get, e.g., how good they are at weighting the information sources appropriately.

Appendices

6.1 Appendix A – Instructions for Experiments 1 and 2 (Chapter 2)

Instructions in Experiment 1:

“We will ask you to play an “octopus” game!

Imagine that you are on a boat and there are 2 octopuses you are trying to find: one is white, and the other one is black. The white octopus has square tentacles and the black octopus has circular tentacles. The 2 octopuses live in different parts of the sea. Sometimes the octopuses will show their tentacles and at other times they will hide at the bottom of the sea.

Your job is to try and figure out where the octopus is!

Once you decide on a location, you can click on the green square (your fishing net), at which point you will see a red dot, which shows you the true location of the octopus on that trial. If the red dot is inside the net, then you correctly guessed the location of the octopus and you get a point!”

Instructions in Experiment 2:

“We will ask you to play an ”octopus” game!

Imagine that you are on a boat and there are 2 octopuses you are trying to find: one is white, and the other one is black. The white octopus has square tentacles and the black octopus has circular tentacles. The 2 octopuses live in different parts of the sea. Sometimes the octopuses will show their tentacles and at other times they will hide at the bottom of the sea.

[It is important to remember that one of the octopuses tends to stay in a particular area, whereas the other one moves quite a bit!]

Your job is to try and figure out where the octopus is!

Once you decide on a location, you can click on the green square (your fishing net), at which point you will see a red dot, which shows you the true location of the octopus on that trial. If the red dot is inside the net, then you correctly guessed the location of the octopus and you get a point!”

6.2 Appendix B – Control Experiment: Likelihood-only task (Chapter 2)

Participants consistently performed sub-optimally across *all* of our experiments. However, when we calculated the optimal weight on the likelihood, we did so under the assumption that people know the true values of the reliability of the sensory cue (i.e., the likelihood). As Sato and Kording (2014) point out, this is clearly not always the case: in fact, in order to perform optimally on our tasks, observers may need to learn about their likelihood variability, as well as prior variability. We, therefore, separately assessed any sensory noise that participants may have had in judging the centroid of the set of dots. If we find that subjects' estimates of the reliability of the likelihood differ from the true values, this would mean that subjects were using incorrect parameters for the task, which may have led to suboptimal performance. We then recomputed the optimal weights based on errors in observers' estimates of centroid location; we could, therefore, test whether subjects were, in fact, near-optimal, when their own sensory variability was taken into account.

Methods

Subjects (*N* = 26; 6 had participated in Experiment 2, 6 had participated in Experiment 3, and the rest had not completed any of the above tasks) were instructed to estimate the centroid of eight dots for different likelihood widths. True locations were drawn from a uniform distribution across the screen (no prior). There were 90 trials overall, with 30 trials of each likelihood width interleaved in a random order. No feedback was given.

For each participant, their error on each trial was calculated by taking the difference between the response and true location for that trial (error = response – true). Their variable error for each likelihood condition was calculated as the standard deviation of the errors. Outliers were excluded prior to calculating the variable error in the same way as described previously.

Results and Discussion of Control Experiment

Participants were significantly worse than ideal (variable error was greater than the true standard deviation of the likelihoods) in the low $(t(25) = 7.45, p < .001)$ and medium $(t(25) = 3.80, p < .001)$, but not the high $(t(25) = 1.48, p = .151)$ variance likelihood conditions (Figure 6.8).

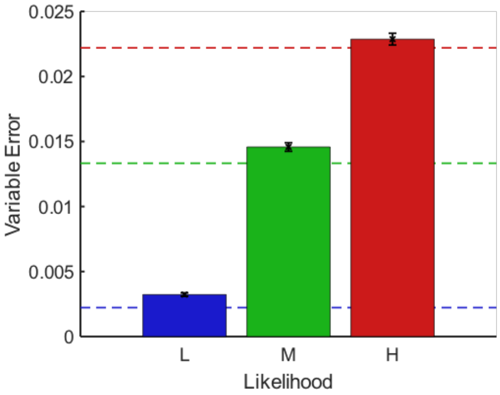

Figure 6.8. Variable error for each likelihood condition in the likelihood-only task. The dashed lines show the true standard deviations of the likelihood in each case (ceiling performance).

This suggests that our optimal predictions place too much weight on the likelihood, as they were calculated based only on the external variability of the sensory cue and failed to also incorporate the added variability from observer's inability to perfectly calculate the dot centroids. We, therefore, recomputed the ideal weight for the likelihood, this time using the measured likelihood variances in the control experiment; we reasoned that this calculation would give us an optimal prediction that better matches our subjects' performance. Our estimates of the likelihood variance increased by 16.66% for the low, 2.96% for the medium and 5.26% for the high likelihood. With such large differences between the true and estimated likelihood variances, we expected that the re-calculated optimal predictions (based on subjects' estimates) will be closer to the observers' data, compared to those based on the true likelihood parameters. We compared these optimal values to subjects' weights in the final block (5) in Experiment 1, and found that they were still significantly different from the empirical data when the variance of the likelihood was high or medium, irrespective of prior variance (all $p < .001$) (see Figure 6.9). No significant differences were observed when the prior variance was wide and the likelihood variance was low $(p = .765)$.

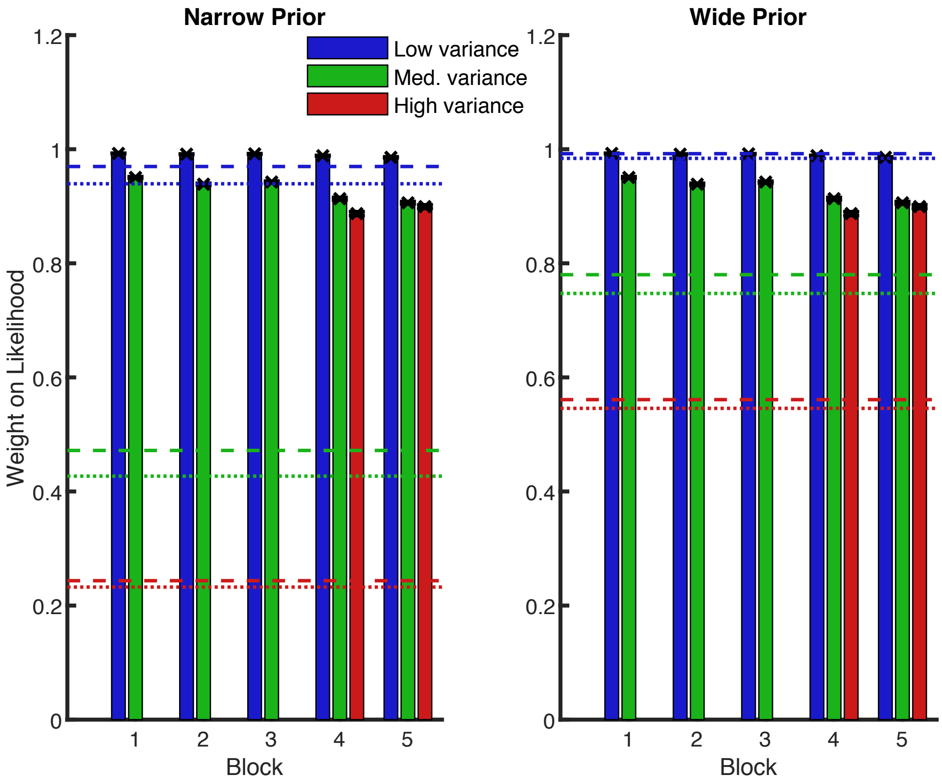

Figure 6.9. Mean weight placed on the likelihood information in each block of Experiment 1. Blue is low variance likelihood, green is medium variance likelihood, red is high variance likelihood. Dashed lines show optimal values. Dotted lines show optimal values, computed using measured likelihood variances in the control experiment. Error bars are +/- 1 SEM.

This pattern of results was surprisingly similar to the one we found when using the predictions of the optimal Bayesian observer, so this analysis did not affect our conclusions on the observers' suboptimal behaviour. In particular, sensory noise in determining the centroid of the "likelihood dots" does not play a major role in explaining subjects' sub-optimality in the task.

6.3 Appendix C (Chapter 2)

The observed lack of statistically significant difference in cue weights does not necessarily imply a lack of substantial difference in terms of performance (points), as previous studies have shown that participants can be "optimally lazy" by deviating from optimal performance in a way that has minimal impact on overall expected score in a task (Acerbi et al., 2017). First off, we computed the optimal response variability σ_b^2 in using both the cue (overall likelihood variability σ_o^2) and the prior as

$$\sigma_b^2 = \frac{\sigma_o^2\,\sigma_p^2}{\sigma_o^2 + \sigma_p^2}$$

Since we are interested in the performance of the model in terms of reward, we then calculated expected gains by first computing the probability of catching an octopus on a given trial as

$$p = P\left(-\frac{w}{2} \leq X \leq \frac{w}{2}\right)$$

where p is the probability that a random draw X from a Gaussian distribution with mean μ (fixed at zero) and standard deviation σ_b^2 will fall within the "hit" distance from the true location, and that distance is half the width of the net $\frac{w}{2}$. The probability of catching the octopus p is then multiplied by the number of trials (per trial type in a block) to calculate the expected number of points.

We then compared expected reward to the average reward earned by those participants who took part in the control experiment and either Experiment 2 or Experiment 3 (*N* = 12; in block 5 only), and found that optimal integration of the sensory cue and prior knowledge (according to participants' overall noise in using the cue) resulted in an expected reward that was higher than what our participants achieved, but only when the variance of the prior was narrow; when the prior variance was wide, they matched quite well; see Figure 6.10. This result is particularly challenging for the notion that people may be "optimally lazy", as this case would result in predicted and obtained reward values being equal. It can be seen that contrary to these predictions, our observers were clearly worse than the optimal observer, and could earn more points when the variance of the prior was narrow; it is, therefore, unlikely that their suboptimal performance could be explained by them being "optimally lazy" (Acerbi et al., 2017).

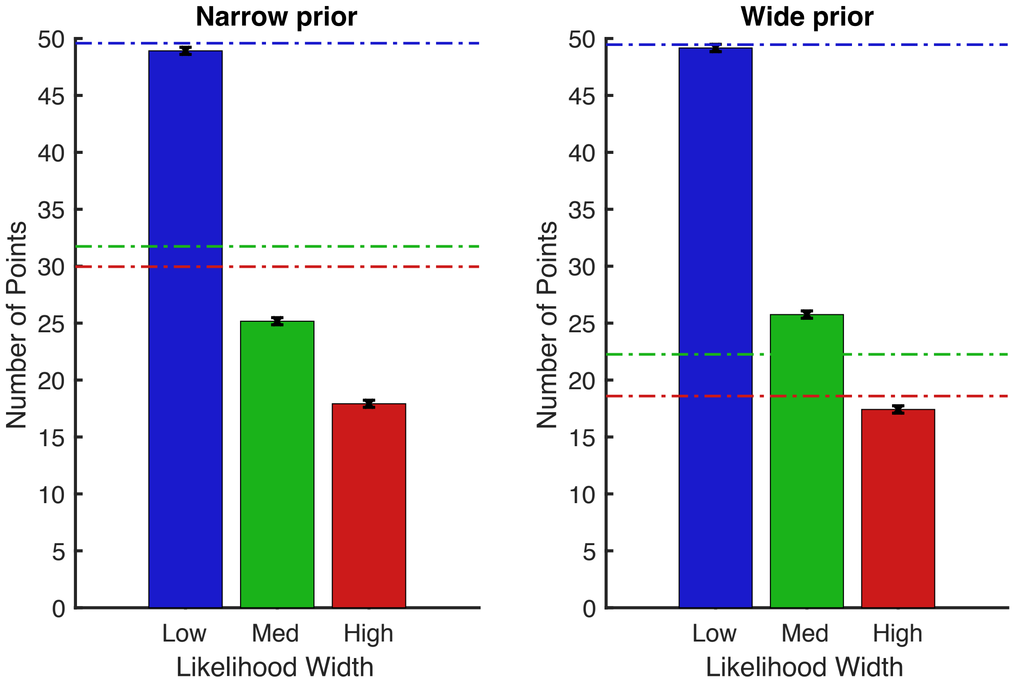

Figure 6.10. Mean number of points earned in Block 5 for participants who took part in Experiment 2 or Experiment 3 and the control task (*N* = 12). Blue is low variance likelihood, green is medium variance likelihood, red is high variance likelihood. Dot-dashed lines show optimal reward values, taking into account participants' overall noise. Error bars are +/- 1 SEM.

References

Abrahamyan, A., Silva, L. L., Dakin, S. C., Carandini, M., & Gardner, J. L. (2016). Adaptable history biases in human perceptual decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(25), E3548–E3557. https://doi.org/10.1073/pnas.1518786113

Acerbi, L., Vijayakumar, S., & Wolpert, D. M. (2014a). On the origins of suboptimality in human probabilistic inference. *PLoS Computational Biology*, *10*(6), e1003661. https://doi.org/10.1371/journal.pcbi.1003661

Acerbi, L., Vijayakumar, S., & Wolpert, D. M. (2014b). On the Origins of Suboptimality in Human Probabilistic Inference. *PLoS Computational Biology*, *10*(6). https://doi.org/10.1371/journal.pcbi.1003661

Acerbi, L., Vijayakumar, S., & Wolpert, D. M. (2017). *Target Uncertainty Mediates Sensorimotor Error Correction*. 1–21. https://doi.org/10.1371/journal.pone.0170466

Acerbi, L., Wolpert, D. M., & Vijayakumar, S. (2012). Internal Representations of Temporal Statistics and Feedback Calibrate Motor-Sensory Interval Timing. *PLoS Computational Biology*, *8*(11). https://doi.org/10.1371/journal.pcbi.1002771

Adams, W. J. (2016). *The Development of Audio-Visual Integration for Temporal Judgements*. 1–17. https://doi.org/10.1371/journal.pcbi.1004865

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

Ahrens, M. B., & Sahani, M. (2011). Observers exploit stochastic models of sensory change to help judge the passage of time. *Current Biology*, *21*(3), 200–206. https://doi.org/10.1016/j.cub.2010.12.043

Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biol*, *14*(3), 257–262.

https://doi.org/10.1016/j.cub.2004.01.029

Alais, D., & Burr, D. (2019). *Cue Combination Within a Bayesian Framework*. 9–31. https://doi.org/10.1007/978-3-030-10461-0_2

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, *77*(6), 1698–1716. https://doi.org/10.1111/j.1467-8624.2006.00968.x

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, *16*(1), 17–42. https://doi.org/10.1007/s11065-006-9002-x

Anderson, M., Bucks, R. S., Bayliss, D. M., & della Sala, S. (2011). Effect of age on dual-task performance in children and adults. *Memory and Cognition*, *39*(7), 1241–1252. https://doi.org/10.3758/s13421-011-0099-7

Andrews, B., Aisenberg, D., D'Avossa, G., & Sapir, A. (2013). Cross-cultural effects on the assumed light source direction: Evidence from English and Hebrew readers. *Journal of Vision*, *13*(13), 1–7. https://doi.org/10.1167/13.13.2

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464–473. https://doi.org/10.1111/j.1467-7687.2009.00937.x

Arffa, S. (2007). The relationship of intelligence to executive function and non-executive function measures in a sample of average, above average, and gifted youth. *Archives of Clinical Neuropsychology*, *22*(8), 969–978. https://doi.org/10.1016/j.acn.2007.08.001

Arnold, D. H., Petrie, K., Murray, C., & Johnston, A. (2019). Suboptimal human multisensory cue combination. *Scientific Reports*, *9*(1), 1–11. https://doi.org/10.1038/s41598-018-37888-7

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

Aston, S., Negen, J., Nardini, M., & Beierholm, U. (2021). Central tendency biases must be accounted for to consistently capture Bayesian cue combination in continuous response data. *BioRxiv*, 2021.03.12.434970.

http://biorxiv.org/content/early/2021/03/12/2021.03.12.434970.abstract

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Baddeley, A. (2007). *Working memory, thought, and action* (Vol. 45). OuP Oxford.

Bagneux, V., Thomassin, N., Gonthier, C., & Roulin, J. L. (2013). Working memory in the processing of the Iowa Gambling Task: An individual differences approach. *PLoS ONE*, *8*(11), 1–6. https://doi.org/10.1371/journal.pone.0081498

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, *20*(7), 1391. https://doi.org/10.1364/josaa.20.001391

Battaglia, P. W., Kersten, D., & Schrater, P. R. (2011). How haptic size sensations improve distance perception. *PLoS Computational Biology*, *7*(6). https://doi.org/10.1371/journal.pcbi.1002080

Baum, G. L., Ciric, R., Roalf, D. R., Betzel, R. F., Moore, T. M., Shinohara, R. T., Kahn, A. E., Vandekar, S. N., Rupert, P. E., Quarmley, M., Cook, P. A., Elliott, M. A., Ruparel, K., Gur, R. E., Gur, R. C., Bassett, D. S., & Satterthwaite, T. D. (2017). Modular Segregation of Structural Brain Networks Supports the Development of Executive Function in Youth. *Current Biology*, *27*(11), 1561-1572.e8. https://doi.org/10.1016/j.cub.2017.04.051

Baum, S. H., Stevenson, R. A., & Wallace, M. T. (2015). Behavioural, Perceptual, and Neural Alterations in Sensory and Multisensory Function in Autism Spectrum Disorder. *Progress in Neurobiology*, *136*, 140–160. https://doi.org/10.1016/j.pneurobio.2015.09.007.Behavioral

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854. https://doi.org/10.1126/science.1158023

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not Noisy, Just Wrong: The Role of Suboptimal Inference in Behavioral Variability. *Neuron*, *74*(1), 30–39. https://doi.org/10.1016/j.neuron.2012.03.016

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Beierholm, U. R., Quartz, S. R., & Shams, L. (2009a). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, *9*(5), 1–9. https://doi.org/10.1167/9.5.23

Beierholm, U. R., Quartz, S. R., & Shams, L. (2009b). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, *9*(2009), 23–23. https://doi.org/10.1167/9.5.23.Introduction

Beierholm, U., Rohe, T., Ferrari, A., Stegle, O., & Noppeney, U. (2020). Using the past to estimate sensory uncertainty. *ELife*, *9*, 1–22. https://doi.org/10.7554/ELIFE.54172

Bejjanki, V. R., & Aslin, R. N. (2020). Optimal, resource-rational or sub-optimal? Insights from cognitive development. *Behavioral and Brain Sciences*, *43*, e4. https://doi.org/DOI: 10.1017/S0140525X19001614

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

Bejjanki, V. R., Randrup, E. R., & Aslin, R. N. (2019). Young children combine sensory cues with learned information in a statistically efficient manner : But task complexity matters. *Developmental Science*, *October*, 1–13. https://doi.org/10.1111/desc.12912

Bentvelzen, A., Leung, J., & Alais, D. (2009). Discriminating audiovisual speed: Optimal integration of speed defaults to probability summation when component reliabilities diverge. *Perception*, *38*(7), 966–987. https://doi.org/10.1068/p6261

Bergen, R S van, & Jehee, J. F. M. (2019). *Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. BioRxiv, 671958*.

Berniker, M., 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

Best, J. R., & Miller, P. H. (2010). A Developmental Perspective on Executive Function. *Child Development*, *81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x.A

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between Executive Function and Academic Achievement from Ages 5 to 17 in a Large, Representative National Sample. *Learning and Individual Differences*, *21*(4), 327–336.

Bianco, R., Harrison, P. M. C., Hu, M., Bolger, C., Picken, S., Pearce, M. T., & Chait, M. (2020). Long-term implicit memory for sequential auditory patterns in humans. *ELife*, *9*, 1–6. https://doi.org/10.7554/eLife.56073

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods*, *17*(3), 399.

Blakemore, M. R., & Snowden, R. J. (1999). The effect of contrast upon perceived speed: A general phenomenon? *Perception*, *28*(1), 33–48. https://doi.org/10.1068/p2722

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-15199-7

Bornstein, A. M., & Daw, N. D. (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience*, *35*(7), 1011–1023. https://doi.org/10.1111/j.1740-9713.2016.00944.xC

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. https://doi.org/10.1037/a0026450

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 433–436.

Braun, A., Urai, A. E., & Donner, T. H. (2018). Adaptive history biases result from confidence-weighted accumulation of past choices. *Journal of Neuroscience*, *38*(10), 2418–2429. https://doi.org/10.1523/JNEUROSCI.2189-17.2017

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*, 127–132. https://doi.org/10.1016/j.cognition.2011.06.010

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

Butler, J. S., Smith, S. T., Campos, J. L., & Bülthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *Journal of Vision*, *10*(11), 1–13. https://doi.org/10.1167/10.11.23

Carandini, M., & Heeger, D. J. (2013). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 1–7. https://doi.org/10.1038/nrn3136.Normalization

Carlson, S. M., Zelazo, P. D., & Faja, S. (2013). *Executive function.*

Cashdollar, N., Ruhnau, P., Weisz, N., & Hasson, U. (2017). The Role of Working Memory in the Probabilistic Inference of Future Sensory Events. *Cerebral Cortex*, *27*(5), 2955–2969. https://doi.org/10.1093/cercor/bhw138

Chalk, M., Seitz, A. R., & Series, P. (2010). Rapidly learned stimulus expectations alter perception of motion. *Journal of Vision*, *10*(8), 2–2. https://doi.org/10.1167/10.8.2

Chambers, C., Gaebler-spira, D., & Kording, K. P. (2018). *The development of Bayesian integration in sensorimotor estimation*. *18*, 1–16.

Champion, R. A., & Adams, W. J. (2007). Modification of the convexity prior but not the light-from-above prior in visual search with shaded objects. *Journal of Vision*, *7*(13), 1–10. https://doi.org/10.1167/7.13.10

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*(7), 287–291. https://doi.org/10.1016/j.tics.2006.05.007

Chen, C. C., & Tyler, C. W. (2015). Shading beats binocular disparity in depth from luminance gradients: Evidence against a maximum likelihood principle for cue combination. *PLoS ONE*, *10*(8), 1–17. https://doi.org/10.1371/journal.pone.0132658

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

Cicchini, Guido Marco, 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, Guido Marco, Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception. *Journal of Vision*, *17*(14), 1–9. https://doi.org/10.1167/17.14.6

Cicchini, Guido Marco, Mikellidou, K., & Burr, D. C. (2018). The functional role of serial dependence. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1890). https://doi.org/10.1098/rspb.2018.1722

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems* (Vol. 105). Springer Science & Business Media.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychological Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.1021/cen-v020n021.p1399

Corbett, J. E., Fischer, J., & Whitney, D. (2011). Facilitating stable representations: Serial dependence in vision. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0016701

Corbin, J. C., Crawford, L. E., & Vavra, D. T. (2017). Misremembering emotion: Inductive category effects for complex emotional stimuli. *Memory and Cognition*, *45*(5), 691–698. https://doi.org/10.3758/s13421-017-0690-7

Corsi, P. M. (1972). Short Title by HUMAN MEMORY AND THE MEDIAL TEMPORAL REGION OF THE BPAIN. *Psychology*.

Crone, E. A., Wendelken, C., Donohue, S., Van Leijenhorst, L., & Bunge, S. A. (2006). Neurocognitive development of the ability to manipulate information in working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(24), 9315–9320. https://doi.org/10.1073/pnas.0510088103

Croydon, A., Karaminis, T., Neil, L., Burr, D., & Pellicano, E. (2017). The light-from-above prior is intact in autistic children. *Journal of Experimental Child Psychology*, *161*, 113–125. https://doi.org/10.1016/j.jecp.2017.04.005

Cui, J. F., Wang, Y., Shi, H. S., Liu, L. L., Chen, X. J., & Chen, Y. H. (2015). Effects of working memory load on uncertain decision-making: Evidence from the Iowa Gambling Task. *Frontiers in Psychology*, *6*(FEB), 1–9. https://doi.org/10.3389/fpsyg.2015.00162

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037–2078.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, *7*(5), 889–904. https://doi.org/10.1162/neco.1995.7.5.889

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, *108*(32), 13341–13346. https://doi.org/10.1073/pnas.1104517108

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Dekker, T., & Lisi, M. (2020). Sensory Development: Integration before Calibration. *Current Biology*, *30*(9), R409–R412. https://doi.org/10.1016/j.cub.2020.02.060

Dekker, T. M., Ban, H., Van Der Velde, B., Sereno, M. I., Welchman, A. E., & Nardini, M. (2015). Late development of cue integration is linked to sensory fusion in cortex. *Current Biology*, *25*(21), 2856–2861. https://doi.org/10.1016/j.cub.2015.09.043

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia*, *37*(10), 1189–1199. https://doi.org/10.1016/S0028-3932(98)00159-6

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dong, D., & Atick, J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems*, *6*(3), 345–358. https://doi.org/10.1088/0954-898x/6/3/003

Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., & Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *ELife*, *3*, 1–19.

https://doi.org/10.7554/elife.03005

Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, L. E. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review*, *17*(2), 224–230. https://doi.org/10.3758/PBR.17.2.224

Duncan, J., Schramm, M., Thompson, R., & Dumontheil, I. (2012). Task rules, working memory, and fluid intelligence. *Psychonomic Bulletin and Review*, *19*(5), 864–870. https://doi.org/10.3758/s13423-012-0225-y

Emberson, L. L., Richards, J. E., & Aslin, R. N. (2015). Top-down modulation in the infant brain: Learning-induced expectations rapidly affect the sensory cortex at 6 months. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(31), 9585–9590. https://doi.org/10.1073/pnas.1510343112

Ernst, M. O. (2007). Learning to combine arbitrary signals from vision and touch. *Journal of Vision*, *7*(2007), 1–14. https://doi.org/10.1167/7.5.7.Introduction

Ernst, M. O. (2008). Multisensory Integration: A Late Bloomer. *Current Biology*, *18*(12), 519–521. https://doi.org/10.1016/j.cub.2008.05.003

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Ester, E. F., Ho, T. C., Brown, S. D., & Serences, J. T. (2014). Variability in visual working memory ability limits the efficiency of perceptual decision making. *Journal of Vision*, *14*(4), 1–12. https://doi.org/10.1167/14.4.2.doi

Fernandes, H. L., Stevenson, I. H., Vilares, I., & Kording, K. P. (2014). The Generalization of Prior Uncertainty during Reaching. *Journal of Neuroscience*, *34*(34), 11470–11484. https://doi.org/10.1523/JNEUROSCI.3882-13.2014

Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic Reweighting of Visual and Vestibular Cues during Self-Motion Perception. *Journal of Neuroscience*, *29*(49), 15601–15612. https://doi.org/10.1523/JNEUROSCI.2574-09.2009

Feulner, B., Postin, D., Schwiedrzik, C. M., & Pooresmaeili, A. (2020). Previous motor actions outweigh sensory information in sensorimotor learning. *BioRxiv*, 2020.12.18.423405.

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

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15822–15826. https://doi.org/10.1073/pnas.232472899

Fornaciai, M., & Park, J. (2018). Attractive Serial Dependence in the Absence of an Explicit Task. *Psychological Science*, *29*(3), 437–446. https://doi.org/10.1177/0956797617737385

Frensch, P. A., & Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition*, *22*(1), 95–110. https://doi.org/10.3758/BF03202765

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite Effects of Recent History on Perception and Decision. *Current Biology*, *27*(4), 590–595. https://doi.org/10.1016/j.cub.2017.01.006

Fritsche, M., Spaak, E., & de Lange, F. P. (2020). A bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *ELife*, *9*, 1–32. https://doi.org/10.7554/eLife.55389

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Gardner, J. L. (2019). Optimality and heuristics in perceptual neuroscience. *Nature Neuroscience*, *22*(4), 514–523.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, *51*(7), 771–781. https://doi.org/10.1016/j.visres.2010.09.027

Geisler, W. S., & Kersten, D. (2002). Illusions, perception and Bayes. *Nature Neuroscience*, *5*(6), 508–510. https://doi.org/10.1038/nn0602-508

Gekas, N., Chalk, M., Seitz, A. R., & Seriès, P. (2013). Complexity and specificity of experimentally induced expectations in motion perception. *BMC Neuroscience*, *14*(Suppl 1), P355. https://doi.org/10.1186/1471-2202-14-S1-P355

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2013). Bayesian data analysis. *Chapman and Hall/CRC*, *146*(4), 165–169. https://doi.org/10.1017/9781316216491.030

Gerardin, P., de Montalembert, M., & Mamassian, P. (2007). Shape from shading: New perspectives from the Polo Mint stimulus. *Journal of Vision*, *7*(11), 1–11. https://doi.org/10.1167/7.11.13

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*. https://doi.org/10.1146/annurev-psych-120709-145346

Glaze, C. M., Kable, J. W., & Gold, J. I. (2015). Normative evidence accumulation in unpredictable environments. *ELife*, *4*(AUGUST2015), 1–27. https://doi.org/10.7554/eLife.08825

Gori, M. (2015). Multisensory integration and calibration in children and adults with and without sensory and motor disabilities. *Multisensory Research*, *28*(1–2), 71–99. https://doi.org/10.1163/22134808-00002478

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young Children Do Not Integrate Visual and Haptic Form Information. *Current Biology*, *18*(9), 694–698. https://doi.org/10.1016/j.cub.2008.04.036

Gori, M., Sandini, G., & Burr, D. (2012). Development of visuo-auditory integration in space and time. *Frontiers in Integrative Neuroscience*, *6*, 77. https://doi.org/10.3389/fnint.2012.00077

Greenhouse, S. W., & Geisser, S. (1959). *ON METHODS IN THE ANALYSIS OF PROFILE variance . Furthermore , an analysis of variance approach permits the analysis of a set of data which cannot be handled by multivariate procedures , namely , the case where n , the number of random vectors , is less t*. *24*(2).

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *352*(1358), 1121–1127. https://doi.org/10.1098/rstb.1997.0095

Gu, S., Satterthwaite, T. D., Medaglia, J. D., Yang, M., Gur, R. E., Gur, R. C., & Bassett, D. S. (2015). Emergence of system roles in normative neurodevelopment. *Proceedings of the National Academy of Sciences of the*

United States of America, *112*(44), 13681–13686. https://doi.org/10.1073/pnas.1502829112

Gur, R. C., Richard, J., Calkins, M. E., Chiavacci, R., Hansen, J. A., Bilker, W. B., Loughead, J., Connolly, J. J., Qiu, H., Mentch, F. D., Abou-Sleiman, P. M., Hakonarson, H., & Gur, R. E. (2012). Age group and sex differences in performance on a computerized neurocognitive battery in children age 8-21. *Neuropsychology*, *26*(2), 251–265. https://doi.org/10.1037/a0026712.Age

Hagmann, P., Sporns, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V. J., Meuli, R., Thiran, J. P., & Grant, P. E. (2010). White matter maturation reshapes structural connectivity in the late developing human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(44), 19067–19072. https://doi.org/10.1073/pnas.1009073107

Haller, S. P. W., Bang, D., Bahrami, B., & Lau, J. Y. F. (2018). Group decision-making is optimal in adolescence. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-33557-x

Hansmann-Roth, S., Kristjánsson, Á., Whitney, D., & Chetverikov, A. (2021). Dissociating implicit and explicit ensemble representations reveals the limits of visual perception and the richness of behavior. *Scientific Reports*, *11*(1), 1–12. https://doi.org/10.1038/s41598-021-83358-y

Harada, T., Tsuruno, M., & Shirokawa, T. (2018). Developmental trajectory of rule management system in children. *Scientific Reports*, *8*(1), 2–12. https://doi.org/10.1038/s41598-018-31235-6

Harrison, L. M., Bestmann, S., Rosa, M. J., Penny, W., & Green, G. G. R. (2011). Time scales of representation in the human brain: Weighing past information to predict future events. *Frontiers in Human Neuroscience*, *5*(APRIL), 1–8. https://doi.org/10.3389/fnhum.2011.00037

Helmholtz, H. (1925). *Treatise on physiological optics. III. The perceptions of vision*.

Heng, J. A., Woodford, M., & Polania, R. (2020). Efficient sampling and noisy decisions. *ELife*, *9*, 1–49. https://doi.org/10.7554/ELIFE.54962

Hewitson, C. L., Sowman, P. F., & Kaplan, D. M. (2018). Interlimb generalization of learned bayesian visuomotor prior occurs in extrinsic coordinates. *ENeuro*, *5*(4), 1–12. https://doi.org/10.1523/ENEURO.0183-18.2018

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: optimal cue combination. *Journal of Vision*, *4*(12), 967–992. https://doi.org/10.1167/4.12.1

Hollingworth, H. L. (1910). The Central Tendency of Judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, *7*(17), 461–469.

Howard, I. S., Wolpert, D. M., & Franklin, D. W. (2013). The effect of contextual cues on the encoding of motor memories. *Journal of Neurophysiology*, *109*(10), 2632–2644. https://doi.org/10.1152/jn.00773.2012

Huang, H., Shu, N., Mishra, V., Jeon, T., Chalak, L., Wang, Z. J., Rollins, N., Gong, G., Cheng, H., Peng, Y., Dong, Q., & He, Y. (2015). Development of human brain structural networks through infancy and childhood. *Cerebral Cortex*, *25*(5), 1389–1404. https://doi.org/10.1093/cercor/bht335

Hudson, T. E., Maloney, L. T., & Landy, M. S. (2008). *Optimal Compensation for Temporal Uncertainty in Movement Planning*. *4*(7). https://doi.org/10.1371/journal.pcbi.1000130

Huizinga, M., & Van Der Molen, M. W. (2007). Age-group differences in set-switching and set-maintenance on the Wisconsin Card Sorting Task. *Developmental Neuropsychology*, *31*(2), 193–215. https://doi.org/10.1080/87565640701190817

Huttenlocher, J., Hedges, L. V, & Vevea, J. L. (2000). *Why do categories affect stimulus judgement.pdf*. *2*, 220–241.

Iliescu, B. F., & Dannemiller, J. L. (2008). *Brain-behavior relationships in early visual development.*

Irwin-Chase, H., & Burns, B. (2000). Developmental Changes in Children's Abilities to Share and Allocate Attention in a Dual Task. *Journal of Experimental Child Psychology*, *77*(1), 61–85. https://doi.org/10.1006/jecp.1999.2557

Ives, A. R. (2015). For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods in Ecology and Evolution*, *6*(7), 828–835. https://doi.org/10.1111/2041-210X.12386

Jacobs, Robert, A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*(24), 4062–4075. https://doi.org/10.1016/S0042-6989(99)00120-0

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive*

Sciences, *6*(8), 345–350. https://doi.org/10.1016/S1364-6613(02)01948-4

Jacobs, R. A., & Fine, I. (1999). Experience-dependent integration of texture and motion cues to depth. *Vision Research*, *39*(24), 4062–4075. https://doi.org/10.1016/S0042-6989(99)00120-0

Jamieson, D. G. (1977). Two presentation order effects. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *31*(4), 184.

Janacsek, K., & Nemeth, D. (2013). Implicit sequence learning and working memory: Correlated or complicated? *Cortex*, *49*(8), 2001–2006. https://doi.org/10.1016/j.cortex.2013.02.012

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

Jazayeri, M., & Shadlen, M. N. (2015). A Neural Mechanism for Sensing and Reproducing a Time Interval. *Current Biology*, *25*(20), 2599–2609. https://doi.org/10.1016/j.cub.2015.08.038

Jones, M., & Love, B. C. (2011). *Bayesian Fundamentalism or Enlightenment ? On the explanatory status and theoretical contributions of Bayesian models of cognition*. 169–231.

Jones, P. R. (2016). A tutorial on cue combination and Signal Detection Theory: Using changes in sensitivity to evaluate how observers integrate sensory information. *Journal of Mathematical Psychology*, *73*, 117–139. https://doi.org/10.1016/j.jmp.2016.04.006

Jones, P. R., & Dekker, T. M. (2017). The development of perceptual averaging: learning what to do, not just how to do it. *Developmental Science*, *April 2017*, e12584. https://doi.org/10.1111/desc.12584

Jost, E., Conway, C. M., Purdy, J. D., & Hendricks, M. A. (2011). *Neurophysiological Correlates of Visual Statistical Learning in Adults and Children*.

Jovanovic, B., & Drewing, K. (2014). The influence of intersensory discrepancy on visuo-haptic integration is similar in 6-year-old children and adults. *Frontiers in Psychology*, *5*(JAN), 1–11. https://doi.org/10.3389/fpsyg.2014.00057

Kalm, K., & Norris, D. (2018). Visual recency bias is explained by a mixture model of internal representations. *Journal of Vision*, *18*(7), 1–15.

https://doi.org/10.1101/228973

Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in Working Capacity as Variation in Executive Attention and Control. In *Variation in working memory* (pp. 21–48).

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

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

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*(3), 321–340. https://doi.org/10.1016/j.cognition.2010.05.011

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. L. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

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

Kersten, Daniel, & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology*, *13*(2), 150–158. https://doi.org/10.1016/S0959-4388(03)00042-4

Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology*, *7*(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8

Kim, T. D., Kabir, M., & Gold, J. I. (2017). Coupled decision processes update and maintain saccadic priors in a dynamic environment. *Journal of Neuroscience*, *37*(13), 3632–3645. https://doi.org/10.1523/JNEUROSCI.3078-16.2017

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy : evidence for a domain general learning mechanism. *Cognition*, *83*, 4–5.

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location , Location , Location : Development of Spatiotemporal Sequence Learning in Infancy. *Child Development*, *78*(5), 1559–1571.

Klahr, D. (1992). Information-processing approaches to cognitive development.pdf. In *Developmental Psychology* (pp. 273–335).

Klahr, D., & Macwhinney, B. (1996). Information Processing. In *Handbook of child psychology* (pp. 631–678).

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(14), 70821. https://doi.org/10.1068/v070821

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01587-5

Knill, D. C., & Pouget, A. (2004). *The Bayesian brain : the role of uncertainty in neural coding and computation*. *27*(12). https://doi.org/10.1016/j.tins.2004.10.007

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian Inference*. Cambridge University Press.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558. https://doi.org/10.1016/S0042-6989(03)00458-9

Kondo, A., Takahashi, K., & Watanabe, K. (2012). Sequential effects in face-attractiveness judgment. *Perception*, *41*(1), 43–49. https://doi.org/10.1068/p7116

Kording, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). *Causal Inference in Multisensory Perception*. *9*. https://doi.org/10.1371/journal.pone.0000943

Kording, K. P., Ku, S., & Wolpert, D. M. (2004). Bayesian Integration in Force Estimation. *Journal of Neurophysiology*, 3161–3165.

Körding, K.P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. https://doi.org/10.1038/nature02169

Körding, Konrad P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10*(7), 319–326. https://doi.org/10.1016/j.tics.2006.05.003

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A. C., Barbosa, L. S., & Gelskov, S. V. (2015). Neural dynamics of prediction and surprise in infants. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms9537

Lak, A., Okun, M., Moss, M., Gurnani, H., Farrell, K., Wells, M. J., Reddy, C. B., Kepecs, A., Harris, K. D., & Carandini, M. (2018). Neural basis of learning guided by sensory confidence and reward value. *BioRxiv*, 1–24. https://doi.org/10.1101/411413

Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A*, *18*(9), 2307. https://doi.org/10.1364/josaa.18.002307

Langer, M. S., & Bulthoff, H. H. (2001). A prior for global convexity in local shape from shading. *Perception*, *30*(4), 403–410.

Laquitaine, S., & Gardner, J. L. (2018). Article A Switching Observer for Human Perceptual A Switching Observer for Human Perceptual Estimation. *Neuron*, 1–13. https://doi.org/10.1016/j.neuron.2017.12.011

Lavín, C., Martín, R. S., & Jubal, E. R. (2014). Pupil dilation signals uncertainty and surprise in a learning gambling task. *Frontiers in Behavioral Neuroscience*, *7*(JAN), 1–8. https://doi.org/10.3389/fnbeh.2013.00218

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

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current Biology*, *24*(21), 2569–2574. https://doi.org/10.1167/18.7.4

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X1900061X

Louw, S., Smeets, J. B. J., & Brenner, E. (2007). Judging surface slant for placing objects: A role for motion parallax. *Experimental Brain Research*, *183*(2), 149–

158. https://doi.org/10.1007/s00221-007-1043-8

Luce, R. D. (1995). Four tensions concerning mathematical modeling in psychology. *Annual Review of Psychology*, *46*(1), 1–27.

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia*, *36*(3), 273–293. https://doi.org/10.1016/S0028-3932(97)00109-7

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*(5), 1357–1372. https://doi.org/10.1111/j.1467-8624.2004.00745.x

Luna, B., Marek, S., Larsen, B., Tervo-Clemmens, B., & Chahal, R. (2015). An Integrative Model of the Maturation of Cognitive Control. *Annual Review of Neuroscience*, *38*, 151–170. https://doi.org/10.1146/annurev-neuro-071714-034054.An

Ma, W. J. (2012). Organizing probabilistic models of perception. *Trends in Cognitive Sciences*, *16*(10), 511–518. https://doi.org/10.1016/j.tics.2012.08.010

Ma, W. J. (2019). Bayesian Decision Models: A Primer. *Neuron*, *104*(1), 164–175. https://doi.org/10.1016/j.neuron.2019.09.037

Ma, W. J., & Huang, W. (2009). *No capacity limit in attentional tracking : Evidence for probabilistic inference under a resource constraint*. *9*(2009), 1–30. https://doi.org/10.1167/9.11.3.Introduction

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual Review of Neuroscience*, *37*, 205–220. https://doi.org/10.1146/annurev-neuro-071013-014017

Maiworm, M., & Röder, B. (2011). Suboptimal auditory dominance in audiovisual integration of temporal cues. *Tsinghua Science and Technology*, *16*(2), 121–132. https://doi.org/10.1016/S1007-0214(11)70019-0

Maloney, L. T. (2002). *Statistical decision theory and biological vision.*

Maloney, L. T., & Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual Neuroscience*, *26*(01), 147. https://doi.org/10.1017/S0952523808080905

Mamassian, P., Landy, M., & Laurence, T. (2002). *Probabilistic Models of the Brain: Perception and Neural Function*.

Mamassian, P., & Landy, M. S. (2001). Interaction of visual prior constraints. *Vision Research*, *41*(20), 2653–2668. https://doi.org/10.1016/S0042-6989(01)00147-X

Manassi, M., Liberman, A., Kosovicheva, A., Zhang, K., & Whitney, D. (2018). Serial dependence in position occurs at the time of perception. *Psychonomic Bulletin and Review*, *25*(6), 2245–2253. https://doi.org/10.3758/s13423-018-1454-5

Manning, C., Jones, P. R., Dekker, T. M., & Pellicano, E. (2018). Psychophysics with children: Investigating the effects of attentional lapses on threshold estimates. *Attention, Perception, and Psychophysics*, *80*(5), 1311–1324. https://doi.org/10.3758/s13414-018-1510-2

Manning, C., Wagenmakers, E.-J., Norcia, M. A., Scerif, G., & Boehm, U. (2020). Perceptual Decision-Making in Children : Age-Related Differences and EEG Correlates. *Computational Brain & Brain*.

Mansouri, F. A., Koechlin, E., Rosa, M. G. P., & Buckley, M. J. (2017). Managing competing goals - A key role for the frontopolar cortex. *Nature Reviews Neuroscience*, *18*(11), 645–657. https://doi.org/10.1038/nrn.2017.111

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information, henry holt and co. *Inc., New York, NY*, *2*(4.2).

Martino, B. De, Bobadilla-suarez, S., Nouguchi, T., Sharot, T., & Love, X. C. (2017). *Social Information Is Integrated into Value and Confidence Judgments According to Its Reliability*. *37*(25), 6066–6074. https://doi.org/10.1523/JNEUROSCI.3880-16.2017

Mattar, M. G., Kahn, D. A., Thompson-Schill, S. L., & Aguirre, G. K. (2016). Varying Timescales of Stimulus Integration Unite Neural Adaptation and Prototype Formation. *Current Biology*, *26*(13), 1669–1676. https://doi.org/10.1016/j.cub.2016.04.065

Maybery, M., Margaret, T., & O’Brien-Malone, A. (1995). Implicit Learning: Sensitive to Age But Not IQ. *Australian Journal of Psuychology*, *47*(1), 8–17.

Meijer, D., Veselič, S., Calafiore, C., & Noppeney, U. (2019). Integration of audiovisual spatial signals is not consistent with maximum likelihood estimation. *Cortex*, *119*, 74–88. https://doi.org/10.1016/j.cortex.2019.03.026

Mendonça, A. G., Drugowitsch, J., Vicente, M. I., DeWitt, E. E. J., Pouget, A., & Mainen, Z. F. (2020). The impact of learning on perceptual decisions and its implication for speed-accuracy tradeoffs. *Nature Communications*, *11*(1), 1–15. https://doi.org/10.1038/s41467-020-16196-7

Miller, R. G. (1986). Beyond ANOVA: Basics of Applied Statistics John Wiley. *Sons, New York, New York*.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*(September), 1–9. https://doi.org/10.3389/fpsyg.2010.00031

Miyazaki, M., Nozaki, D., & Nakajima, Y. (2005). Testing Bayesian Models of Human Coincidence Timing. *Journal of Neurophysiology*, *94*(1), 395–399. https://doi.org/10.1152/jn.01168.2004.

Morcos, A. S., & Harvey, C. D. (2016). History-dependent variability in population dynamics during evidence accumulation in cortex. *Nature Neuroscience*, *19*(12), 1672–1681. https://doi.org/10.1038/nn.4403.History-dependent

Morgenstern, Y., Murray, R. F., & Harris, L. R. (2011). The human visual system’s assumption that light comes from above is weak. *Proceedings of the National Academy of Sciences*, *108*(30), 12551–12553. https://doi.org/10.1073/pnas.1100794108

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods*, *222*, 250–259. https://doi.org/10.1016/j.jneumeth.2013.10.024

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(39), 17041–17046. https://doi.org/10.1073/pnas.1001699107

Nardini, M., Begus, K., & Mareschal, D. (2013). Multisensory uncertainty reduction for hand localization in children and adults. *Journal of Experimental Psychology. Human Perception and Performance*, *39*(3), 773–787. https://doi.org/10.1037/a0030719

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of Cue Integration in Human Navigation. *Current Biology*, *18*(9), 689–693. https://doi.org/10.1016/j.cub.2008.04.021

Negen, J., Chere, B., Bird, L. A., Taylor, E., Roome, H. E., Keenaghan, S., Thaler, L., & Nardini, M. (2019). Sensory cue combination in children under 10 years of age. *Cognition*, *193*(June). https://doi.org/10.1016/j.cognition.2019.104014

Negen, J., Wen, L., Thaler, L., & Nardini, M. (2018). Bayes-Like Integration of a New Sensory Skill with Vision. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-35046-7

Nemeth, D., Janacsek, K., & Fiser, J. (2013). Age-dependent and coordinated shift in performance between implicit and explicit skill learning. *Frontiers in Computational Neuroscience*, *7*(OCT), 1–13. https://doi.org/10.3389/fncom.2013.00147

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective and Behavioral Neuroscience*, *12*(2), 241–268. https://doi.org/10.3758/s13415-011-0083-5

Norton, E. H., Acerbi, L., Ma, W. J., & Landy, M. S. (2018). *Human online adaptation to changes in prior probability Author summary*. 1–24.

Norton, E. H., Acerbi, L., Ma, W. J., & Landy, M. S. (2019). Human online adaptation to changes in prior probability. *PLoS Computational Biology*, *15*(7), 1–26. https://doi.org/10.1101/483842

O'Hara, R. B. (2009). How to make models add up - a primer on GLMMs. *Annales Zoologici Fennici*, *46*(2), 124–137. https://doi.org/10.5735/086.046.0205

Ohshiro, T., Angelaki, D. E., & Deangelis, G. C. (2011). A normalization model of multisensory integration. *Nature Neuroscience*, *14*(6), 775–782. https://doi.org/10.1038/nn.2815

Olkkonen, M., & Allred, S. R. (2014). Short-term memory affects color perception in context. *PLoS ONE*, *9*(1), 1–11. https://doi.org/10.1371/journal.pone.0086488

Olkkonen, M., Mccarthy, P. F., & Allred, S. R. (2014). *The central tendency bias in color perception : Effects of internal and external noise*. *14*(2014), 1–15. https://doi.org/10.1167/14.11.5.doi

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23), 2451–2468. https://doi.org/10.1016/S0042-6989(03)00435-8

Papadimitriou, C., Ferdoash, A., & Snyder, L. H. (2015). Ghosts in the machine: Memory interference from the previous trial. *Journal of Neurophysiology*, *113*(2), 567–577. https://doi.org/10.1152/jn.00402.2014

Park, J., Yoon, H. D., Yoo, T., Shin, M., & Jeon, H. A. (2020). Potential and efficiency of statistical learning closely intertwined with individuals' executive functions: a mathematical modeling study. *Scientific Reports*, *10*(1), 1–13. https://doi.org/10.1038/s41598-020-75157-8

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Pereg, M., & Meiran, N. (2019). Rapid instructed task learning (but not automatic effects of instructions) is influenced by working memory load. *PLoS ONE*, *14*(6), 1–25. https://doi.org/10.1371/journal.pone.0217681

Petrini, K., Caradonna, A., Foster, C., Burgess, N., & Nardini, M. (2016). How vision and self-motion combine or compete during path reproduction changes with age. *Scientific Reports*, *6*, 1–10. https://doi.org/10.1038/srep29163

Petrini, K., Remark, A., Smith, L., & Nardini, M. (2014). When vision is not an option: Children's integration of auditory and haptic information is suboptimal. *Developmental Science*, *17*(3), 376–387. https://doi.org/10.1111/desc.12127

Petzschner, F. H., Maier, P., & Glasauer, S. (2012). Combining Symbolic Cues with Sensory Input and Prior Experience in an Iterative Bayesian Framework. *Frontiers in Integrative Neuroscience*, *6*(August), 1–18. https://doi.org/10.3389/fnint.2012.00058

Pickard-Jones, B., d'Avossa, G., & Sapir, A. (2020). 3D shape-from-shading relies on a light source prior that does not change with age. *Vision Research*, *177*(September), 88–96. https://doi.org/10.1016/j.visres.2020.09.002

Power, J. D., Fair, D. A., Schlaggar, B. L., & Petersen, S. E. (2010). The

Development of Human Functional Brain Networks. *Neuron*, *67*(5), 735–748. https://doi.org/10.1016/j.neuron.2010.08.017

Pretto, P., Bresciani, J. P., Rainer, G., & Bülthoff, H. H. (2012). Foggy perception slows us down. *ELife*, *2012*(1), 1–12. https://doi.org/10.7554/eLife.00031

Preuschoff, K., ’t Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline’s role in decision making. *Frontiers in Neuroscience*, *5*(SEP), 1–12. https://doi.org/10.3389/fnins.2011.00115

Prsa, M., Gale, S., & Blanke, O. (2012). Self-motion leads to mandatory cue fusion across sensory modalities. *Journal of Neurophysiology*, *108*(8), 2282–2291. https://doi.org/10.1152/jn.00439.2012

Qamar, A. T., Cotton, R. J., George, R. G., Beck, J. M., Prezhdo, E., Laudano, A., Tolias, A. S., & Ma, W. J. (2013). Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(50), 20332–20337. https://doi.org/10.1073/pnas.1219756110

Rahnev, D., & Denison, R. N. (2018). *Suboptimality in perceptual decision making*. https://doi.org/10.1017/S0140525X18000936

Ratcliff, R., Love, J., Thompson, C. A., & Opfer, J. E. (2012). *Children Are Not Like Older Adults : A Diffusion Model Analysis of Developmental Changes in Speeded Responses*. *83*(1), 367–381. https://doi.org/10.1111/j.1467-8624.2011.01683.x

Reynolds, J. H., & Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*, *61*(2), 168–185. https://doi.org/10.1016/j.neuron.2009.01.002

Riskey, D. R., Parducci, A., & Beauchamp, G. K. (1979). Effects of context in judgments of sweetness and pleasantness. *Perception & Psychophysics*, *26*(3), 171–176. https://doi.org/10.3758/BF03199865

Roberson, D., Damjanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a “category adjustment” model. *Memory and Cognition*, *35*(7), 1814–1829. https://doi.org/10.3758/BF03193512

Rohlf, S., Habets, B., von Frieling, M., & Röder, B. (2017). Infants are superior in implicit crossmodal learning and use other learning mechanisms than adults. *ELife*, *6*, 1–23. https://doi.org/10.7554/eLife.28166

Rohlf, S., Li, L., Bruns, P., & Röder, B. (2020). Multisensory Integration Develops Prior to Crossmodal Recalibration. *Current Biology*, *30*(9), 1726-1732.e7. https://doi.org/10.1016/j.cub.2020.02.048

Romo, R., Brody, C. D., Hernández, A., & Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, *399*(6735), 470–473. https://doi.org/10.1038/20939

Rosas, P., Wagemans, J., Ernst, M. O., & Wichmann, F. A. (2005). Texture and haptic cues in slant discrimination: reliability-based cue weighting without statistically optimal cue combination. *Journal of the Optical Society of America A*, *22*(5), 801. https://doi.org/10.1364/josaa.22.000801

Rosenberg, A., Patterson, J. S., & Angelaki, D. E. (2015). A computational perspective on autism. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(30), 9158–9165. https://doi.org/10.1073/pnas.1510583112

Rubenstein, J. L. R., & Merzenich, M. M. (2003). Model of autism: Increased ratio of excitation/inhibition in key neural systems. *Genes, Brain and Behavior*, *2*(5), 255–267. https://doi.org/10.1034/j.1601-183X.2003.00037.x

Ryan, L. J. (2011). Temporal context affects duration reproduction. *Journal of Cognitive Psychology*, *23*(1), 157–170. https://doi.org/10.1080/20445911.2011.477812

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Sato, Y., & Kording, K. P. (2014). How much to trust the senses: Likelihood learning. *Journal of Vision*, *14*(13), 13–13. https://doi.org/10.1167/14.13.13

Sato, Yoshiyuki, & Kording, K. P. (2014). How much to trust the senses: Likelihood learning. *Journal of Vision*, *14*(13), 1–13. https://doi.org/10.1167/14.13.13.doi

Satterthwaite, T. D., Wolf, D. H., Erus, G., Ruparel, K., Elliott, M. A., Gennatas, E. D., Hopson, R., Jackson, C., Prabhakaran, K., Bilker, W. B., Calkins, M. E., Loughead, J., Smith, A., Roalf, D. R., Hakonarson, H., Verma, R., Davatzikos, C., Gur, R. C., & Gur, R. E. (2013). Functional maturation of the executive system during adolescence. *Journal of Neuroscience*, *33*(41), 16249–16261. https://doi.org/10.1523/JNEUROSCI.2345-13.2013

Scarfe, P., & Glennerster, A. (2018). Experimentally disambiguating models of sensory cue combination. *Journal of Vision*, *18*(10), 788.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*(3), 414.

Schott, B. H., Wüstenberg, T., Lücke, E., Pohl, I. M., Richter, A., Seidenbecher, C. I., Pollmann, S., Kizilirmak, J. M., & Richardson-Klavehn, A. (2019). Gradual acquisition of visuospatial associative memory representations via the dorsal precuneus. *Human Brain Mapping*, *40*(5), 1554–1570. https://doi.org/10.1002/hbm.24467

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature Reviews Neuroscience*, *8*(7), 522–535. https://doi.org/10.1038/nrn2155

Sciutti, A., Burr, D., Saracco, A., Sandini, G., & Gori, M. (2014). Development of context dependency in human space perception. *Experimental Brain Research*, *232*(12), 3965–3976. https://doi.org/10.1007/s00221-014-4021-y

Senna, I., Parise, C. V., & Ernst, M. O. (2015). Hearing in slow-motion: Humans underestimate the speed of moving sounds. *Scientific Reports*, *5*, 1–5. https://doi.org/10.1038/srep14054

Seydell, A. (2010). Adapting internal statistical models for interpreting visual cues to depth. *Journal of Vision*, *10*(4), 1–27. https://doi.org/10.1167/10.4.1

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*(4), 1916–1936. https://doi.org/10.1152/jn.2001.86.4.1916

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, *14*(9), 425–432. https://doi.org/10.1016/j.tics.2010.07.001

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and*

Language, *81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001.Statistical

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural Image Statistics and Neural Representation. *Annual Review of Neuroscience*, *24*(1), 1193–1216.

Snowden, R. J., Stimpson, N., & Ruddle, R. A. (1998). Speed perception fogs up as visibility drops. *Nature*, *392*(6675), 450–450.

Sotiropoulos, G., Seitz, A. R., & Seris, P. (2011). Changing expectations about speed alters perceived motion direction. *Current Biology*, *21*(21), R883–R884. https://doi.org/10.1016/j.cub.2011.09.013

Spicer, J., Sanborn, A. N., & Beierholm, U. R. (2020). Using Occam's razor and Bayesian modelling to compare discrete and continuous representations in numerosity judgements. *Cognitive Psychology*, *122*(July), 101309. https://doi.org/10.1016/j.cogpsych.2020.101309

Stine, G. M., Zylberberg, A., Ditterich, J., & Shadlen, M. N. (2020). Differentiating between integration and non-integration strategies in perceptual decision making. *ELife*, 1–28. https://doi.org/10.1101/2020.01.24.918169

Stocker, A. 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

Stone, L. S., & Thompson, P. (1992). Human Speed Perception is Contrast Dependent. *Vision Research*, *32*(8), 1535–1549.

Stone, J. V., & Pascalis, O. (2010). Footprints sticking out of the sand. Part 1: Children's perception of naturalistic and embossed symbol stimuli. *Perception*, *39*(9), 1254–1260. https://doi.org/10.1068/p6725

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Sun, Jennifer, & Perona, P. (1997). Shading and stereo in early perception of shape and reflectance. *Perception*, *26*(4), 519–529.

Sun, Jingwei, Li, J., & Zhang, H. (2019). Human representation of multimodal distributions as clusters of samples. *PLoS Computational Biology*, *15*(5), 1–29. https://doi.org/10.1371/journal.pcbi.1007047

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008

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

Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*, *6*, 2–8. https://doi.org/10.1038/srep32239

Taubert, J., Van Der Burg, E., & Alais, D. (2016). Love at second sight: Sequential dependence of facial attractiveness in an on-line dating paradigm. *Scientific Reports*, *6*(March), 2–6. https://doi.org/10.1038/srep22740

Thomas, R., Nardini, M., & Mareschal, D. (2010). Interactions between “light-from-above” and convexity priors in visual development. *Journal of Vision*, *10*(8), 6.1-7. https://doi.org/10.1167/10.8.6

Thomason, Moriah, E., Race, E., Burrows, B., Whitfield-Gabrieli, S., Glover, G. H., & Gabrieli, J. D. E. (2009). Development of Spatial and Verbal Working Memory Capacity in the Human Brain. *Journal of Cognitive Neiuroscience*, *21*(2), 316–332. https://doi.org/10.1162/jocn.2008.21028.Development

Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research*, *22*(3), 377–380. https://doi.org/10.1016/0042-6989(82)90153-5

Trommershauser, J., Kording, K., & Landy, M. S. (2011). *Sensory cue integration*. Oxford University Press.

Tummeltshammer, K., Amso, D., French, R. M., & Kirkham, N. Z. (2017). Across space and time: infants learn from backward and forward visual statistics. *Developmental Science*, *20*(5), 1–17. https://doi.org/10.1111/desc.12474

Tummeltshammer, K. S., & Kirkham, N. Z. (2013). Learning to look: Probabilistic variation and noise guide infants’ eye movements. *Developmental Science*, *16*(5), 760–771. https://doi.org/10.1111/desc.12064

Urai, A. E., de Gee, J. W., Tsetsos, K., & Donner, T. H. (2019). Choice History Biases Subsequent Evidence Accumulation. *ELife*, 1–34. https://doi.org/10.1101/251595

Usher, M., & McClelland, J. L. (2001). The Time Course of Perceptual Choice: The

Leaky, Competing Accumulator Model. *Psychological Review*, *108*(3), 550–592.

van Beers, R., Sittig, A., & van der Gon, D. (1999). An Integration of Proprioceptive and Visual Position-Information: An Experimentally Supported Model. *Integration of Proprioceptive and Visual Position-Information: An Experimentally Supported Model*, *3*(81), 1355–1364.

van Bergen, Ruben S., & Jehee, J. F. M. (2019). Probabilistic Representation in Human Visual Cortex Reflects Uncertainty in Serial Decisions. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *39*(41), 8164–8176. https://doi.org/10.1523/JNEUROSCI.3212-18.2019

Van de Cruys, S., Vanmarcke, S., Steyaert, J., & Wagemans, J. (2018). Intact perceptual bias in autism contradicts the decreased normalization model. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-31042-z

Van Den Berg, R., & Ma, W. J. (2012). Robust averaging during perceptual judgment is not optimal. *Proceedings of the National Academy of Sciences*, *109*(13), E736–E736. https://doi.org/10.1073/pnas.1119078109

Vilares, I., Howard, J. D., Fernandes, H. L., Gottfried, J. A., & Kording, K. P. (2012). Differential representations of prior and likelihood uncertainty in the human brain. *Current Biology*, *22*(18), 1641–1648. https://doi.org/10.1016/j.cub.2012.07.010

Vilares, I., & Kording, K. (2011). Bayesian models: The structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*(1), 22–39. https://doi.org/10.1111/j.1749-6632.2011.05965.x

Virag, M., Janacsek, K., Horvath, A., Bujdoso, Z., Fabo, D., & Nemeth, D. (2015). Competition between frontal lobe functions and implicit sequence learning: evidence from the long-term effects of alcohol. *Experimental Brain Research*, *233*(7), 2081–2089.

Warton, D. I., Lyons, M., Stoklosa, J., & Ives, A. R. (2016). Three points to consider when choosing a LM or GLM test for count data. *Methods in Ecology and Evolution*, *7*(8), 882–890. https://doi.org/10.1111/2041-210X.12552

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, and Computers*, *35*(4), 550–564. https://doi.org/10.3758/BF03195534

Wechsler, D. (2011). Wechsler Abbreviated Scale of Intelligence Second Edition (WASI-II) San Antonio. *TX: Pearson.[Google Scholar]*.

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

Welchman, A. E., Lam, J. M., & Bülthoff, H. H. (2008). Bayesian motion estimation accounts for a surprising bias in 3D vision. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(33), 12087–12092. https://doi.org/10.1073/pnas.0804378105

Wilde, N. J., Strauss, E., & Tulsky, D. S. (2004). Memory span on the Wechsler scales. *Journal of Clinical and Experimental Neuropsychology*, *26*(4), 539–549.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135. https://doi.org/10.1167/4.12.11

Wilson, R. C., & Collins, A. G. E. (2019). Ten simple rules for the computational modeling of behavioral data. *ELife*, *8*, 1–33. https://doi.org/10.7554/eLife.49547

Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, *12*(12). https://doi.org/10.1038/nrn3112

Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants Learn About Objects From Statistics and People. *Developmental Psychology*, *47*(5), 1220–1229. https://doi.org/10.1037/a0024023

Xia, Y., Leib, A. Y., & Whitney, D. (2016). Serial dependence in the perception of visual variance. *Journal of Vision*, *16*(5), 1–8. https://doi.org/10.1167/18.7.4

Yin, C., Wang, H., Wei, K., & Körding, K. P. (2019). Sensorimotor priors are effector dependent. *Journal of Neurophysiology*, *122*(1), 389–397. https://doi.org/10.1152/jn.00228.2018

Yu, A. J., & Cohen, J. D. (2009). Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference*, 1873–1880.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308. https://doi.org/10.1016/j.tics.2006.05.002

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). *The development of executive function in childhood.*

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*(1), 37–63. https://doi.org/10.1016/S0885-2014(96)90027-1

Zhou, Y., Acerbi, L., & Ma, W. J. (2018). *The Role of Sensory Uncertainty in Simple Perceptual Organization Author summary*.