THE UNIVERSITY
of EDINBURGH

Perceptual Bayesian Inference in Autism and Schizophrenia

Povilas Karvelis

Doctor of Philosophy

Institute for Adaptive and Neural Computation
School of Informatics
University of Edinburgh

2020

Abstract

Recent theories in the field of computational psychiatry regard schizophrenia (SCZ) and autistic spectrum disorders (ASD) as impairments in Bayesian inference performed by the brain. In Bayesian terms, perception is a result of optimal real-time integration of sensory information ('likelihood'), which is intrinsically noisy and ambiguous, and prior expectations about the states of the world ('prior'), which serve to disambiguate the meaning of the sensory information. Priors capture statistical regularities in the environment and are constantly updated to keep up with any changes in these regularities. The extent to which prior or likelihood dominate perception depends on the uncertainty with which they are represented, with less uncertainty resulting in more influence. Individuals with ASD and SCZ might show impairments in how they update their priors and/or how much uncertainty there is ascribed to prior and likelihood representations, leading to differences in inference. While this Bayesian account can be argued to be consistent with many previous experimental findings and symptoms of SCZ and ASD, recent experimental work inspired by these ideas has produced mixed results.

In this work, we investigated possible Bayesian impairments in SCZ and ASD experimentally by addressing some of the methodological limitations of the previous work. Most notably, we used an experimental design that allows to disentangle and quantify separate influences of priors and likelihoods, and we tested both SCZ and ASD patient groups as well as autistic and schizotypy traits in the general population. We administered a visual motion perception task that rapidly induces prior expectations about the stimulus motion direction, leading to biases and occasional hallucinations that can be well described by a Bayesian model. In this task, autistic traits were found to be associated with reduced biases, which was underlied by more precise sensory representations, while the acquired priors were not affected by autistic traits. Patients with ASD, however, showed no evidence of increased sensory precision, while there also were no impairments in the acquisition of priors. We also found no effects in the acquisition of priors or sensory representations along schizotypy traits and in patients with SCZ. However, under conditions of high ambiguity SCZ patients were less likely to hallucinate the stimulus than controls.

The second part of the thesis is focused on further exploratory analyses conducted using these same datasets. First, we investigated post-perceptual repulsion effects in our task and whether they were related to trait or group differences. We found clear evidence of repulsion from the cardinal directions. In addition to that, we found evidence for a repulsion from the central reference angle, which was randomly selected for each participant and which could only be inferred from the stimulus statistics. Furthermore, we found the repulsion from the central reference angle to be reduced along schizotypy traits. Interestingly, in both SCZ and ASD groups this repulsion was also found to be negligible. While these results are exploratory, they might point to a trans-diagnostic features of ASD and SCZ. Second, we investigated within-trial dynamics of evidence accumulation by constructing a Continuous Choice Drift Diffusion Model (CDM) – an extension of the classical binary choice drift diffusion model. The results of this model showed that increased sensory precision along AQ found in a Bayesian model was underlied by faster drift rates, while slower responses and reduced hallucinations in SCZ were explained by a larger decision threshold. In addition, this model provided a more complete characterization of the performance in this task (by including reaction times) and it serves to emphasize the importance of accounting for exposure to stimulus duration and judgement time in future studies investigating Bayesian inference.

Together, this work provides novel experimental evidence that speaks to the hypothesis of impaired Bayesian inference in ASD and SCZ. Furthermore, the analysis of reference repulsion effects and within-trial dynamics provide additional insight related to SCZ and ASD differences that extend beyond the Bayesian framework.

Lay summary

Mental disorders are not well understood. The field of computational psychiatry aims to improve our understanding of mental disorders by studying them through the prism of normative computational models of the brain. One of the fundamental challenges for the brain is to infer its surroundings from the limited and noisy sensory information. To perform this inference, the brain has to rely on accumulated knowledge and expectations about the environment. Such knowledge often derives from statistical regularities in the environment. It has been proposed that the way people with schizophrenia (SCZ) and autism spectrum disorders (ASD) accumulate such knowledge and how they integrate it with sensory information is imbalanced, forming the basis of these disorders.

In this thesis, we perform experiments to investigate this hypothesis. We use a computerized visual motion estimation task, where participants have to estimate the direction of motion of ambiguous stimuli. Unbeknownst to participants, two of the motion directions are presented more frequently, inducing subconscious expectations that lead to biases in perception. Using this task, we study people with SCZ and ASD diagnosis as well as schizotypy and autistic traits in the general population. Across all of these groups we find no impairments in the acquisition of statistical knowledge about the stimulus distribution. We nonetheless find increased precision in sensory representations along autistic traits, which leads to reduced biases in our task. However, in patients diagnosed with ASD we find no evidence of increased sensory precision. Moreover, while patients with SCZ showed similar biases in estimating motion direction of the stimulus, they were less likely to falsely perceive it when it was not presented. Our modelling analysis suggested this was due to patients being more conservative and waiting longer before making a response on each trial. Taken together, this work provides empirical results and methodological ideas that can help further refine the computational accounts of SCZ and ASD disorders.

Acknowledgements

I would like to thank Peggy Seriès for supporting me and guiding me through my PhD (and for encouraging me to undertake such venture in the first place!). I thank Stephen Lawrie and Katie Richards for their excellent work with recruiting and testing patients, as well as for sharing their first-hand experience of the complexities of mental disorders. I thank Aaron Seitz and Russell Cohen Hoffing for all the interesting and challenging collaborations. A special thanks goes to my fellow lab members Samuel Rupprechter and James Raymond for all the moments of frustration and disappointment that we have shared and for our short-lived but high-yielding get-togethers for discussing statistical and computational methods. I also thank Chris Lucas and Matthias Henning for their insightful feedback during my yearly progress reviews. I cheerfully thank Adam Linson for his admirable, refreshing and contagious interest in ideas, and for the inspiration for the continuous choice drift diffusion model. Finally, I would like to thank myself for being there though it all - I cannot imagine how I could have done this without myself, so thanks.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Povilas Karvelis

Table of Contents

1

General Introduction

"It appears that things are not as they appear"

– Unknown

1.1 Computational psychiatry

Modern-day psychiatry is far from precise. Current diagnostic categories are mostly based on surface symptomatology (DSM-V, APA, 2013; ICD-10, WHO, 1993), and lack validity (e.g., Kendell and Jablensky, 2003; Dalal and Sivakumar, 2009), while treatments often rely on trial and error (e.g., McMahon, 2014) and focus on managing symptoms rather than treating the underlying illness (e.g., Jacob, 2015). That is primarily a consequence of not having a sufficient understanding of how the observed mental dysfunctions arise from the complex computations performed by the brain, and not knowing what exactly among these computations goes wrong (e.g., Hyman, 2007). Computational psychiatry steps in to fill this gap: rooted in computational neuroscience it aims to understand mental disorders in terms of normative computational accounts of brain function and neurobiology (Adams et al., 2016; Stephan and Mathys, 2014; Montague et al., 2012; Huys et al., 2011). Such computational models are used together

with experimental data to study how disruption at different levels of information processing propagate upward to produce dysfunction at the level of behavior. Importantly, explicit implementations of different models allows for quantitative and rigorous model comparison for testing alternative hypotheses about brain function and dysfunction. In addition to uncovering the mechanisms of mental disorders (*theory-driven approach*), computational psychiatry also employs machine learning tools for finding patterns in large amounts of data, which could be used for building diagnostic tools that stratify current diagnostic categories and that are more predictive of the treatment outcomes than the conventional methods (*data-driven approach*; Yahata et al., 2017; Huys et al., 2016).

1.2 Perception as Bayesian inference

The fundamental challenge for the brain is to infer states of the world from inherently noisy sensory information. To do it effectively, the brain must use an internal model of the environment (Conant and Ross Ashby, 1970). The Bayesian brain hypothesis formalizes this process as a statistically optimal integration of incoming sensory information (*likelihood*) and knowledge or expectations about the environment (*prior*) (von Helmholtz, 1867; Gregory, 1980; Dayan et al., 1995; Lee and Mumford, 2003; Kersten et al., 2004; Doya, 2007). Here, optimality refers to weighting prior and likelihood proportionally to their associated certainty. That is, the influence of prior increases as the reliability of sensory information decreases and vice versa. In this sense, people have been shown to perform optimally on a variety of perceptual tasks, including multisensory and sensorimotor integration (for reviews, see Knill and Pouget, 2004; Pouget et al., 2013).

Interestingly, the characteristic biases found in many perceptual tasks and illusions can be explained by priors that correspond to natural-scene statistics,

such as expectations for cardinal orientations (Girshick et al., 2011) and global convexity (Langer and Bülthoff, 2001), expectations that objects tend to be static or move slowly (Weiss et al., 2002), that the light comes from above (Sun and Perona, 1998) or that someone's gaze is directed at us (Mareschal et al., 2013). The correspondence between environmental statistics and internal priors is another sense in which people can be considered to be optimal. To remain optimal, however, the priors have to be dynamically updated to take into account any new regularities in the environment. Multiple studies have demonstrated that people can update their priors to approximate the statistics of the stimulus on relatively short time scales (e.g., Adams et al., 2004; Körding and Wolpert, 2004; Sotiropoulos et al., 2011; for a review, see Seriès and Seitz, 2013).

On the neural level, hierarchical Bayesian inference is supported by the brain's hierarchical architecture (the recurrent feedforward/feedback loops) and can be implemented via predictive coding[1] (Srinivasan et al., 1982; Mumford, 1992; Rao and Ballard, 1999; Lee and Mumford, 2003; Friston, 2008; Bastos et al., 2012). In predictive coding, each level in the inference hierarchy predicts the input of the level below it. The mismatch between predictions and inputs (*prediction error*) is sent back to the higher levels and drives the updating of priors/predictions. The prediction errors are weighted by the ratio of sensory likelihood and prior precisions (*learning rate*) to modulate the size of the update: the more precise the likelihood (or the less precise the prior) the larger the update. This process repeats iteratively as the brain continuously strives toward minimization of prediction error, i.e., as it builds ever more accurate model of the world. Importantly, encoding of sensory prediction errors and uncertainty at different hierarchical levels has been associated with specific neuromodulatory systems, providing further

[1] note that while Bayesian inference and predictive coding often go hand in hand, the former primarily deals with uncertainty, while the latter is rooted in redundancy reduction; thus, one does not necessarily imply the other (see Aitchison and Lengyel, 2017)

neurobiological basis for Bayesian inference and learning (Fiorillo et al., 2003; Angela and Dayan, 2005; Iglesias et al., 2013; Friston et al., 2012; Moran et al., 2013; Schwartenbeck et al., 2015; Marshall et al., 2016; Diaconescu et al., 2017; for a review, see Iglesias et al., 2017).

1.3 Bayesian accounts of Schizophrenia and ASD

Capturing such fundamental aspects of brain function, Bayesian inference provides a normative framework for understanding mental disorders. In particular, Bayesian accounts have been proposed for schizophrenia (SCZ), which is primarily characterized by reoccurring hallucinations and delusions (Hemsley and Garety, 1986; Fletcher and Frith, 2009; Corlett et al., 2009; Adams et al., 2013; Jardri and Denève, 2013; Sterzer et al., 2018; Heinz et al., 2019) and autistic spectrum disorder (ASD), which is primarily characterized by severe social interaction difficulties, inflexible behavioral patterns, as well as sensory hypo- and hyper-sensitivities (Pellicano and Burr, 2012b; Van de Cruys et al., 2014; Lawson et al., 2014; Sinha et al., 2014; Rosenberg et al., 2015; Palmer et al., 2017). In both disorders, precision ascribed to sensory evidence and prior expectations is proposed to be imbalanced, resulting in disproportionate influence of either top-down or bottom-up signals. In ASD, the influence of priors is thought to be weaker, while in SCZ both weaker and stronger priors have been proposed to explain different findings.

In SCZ, overly precise and salient sensory information might be underlied by increased tonic dopamine levels at D2 receptors (Howes and Kapur, 2009; Winton-Brown et al., 2014), while weaker influence of top-down priors might result from the hypofunction of feedback-mediating N-methyl-d-aspartate receptors (NMDA-R) (Olney and Farber, 1995; Friston et al., 2016). Failure to accurately predict

sensory experiences could explain hallucinations (for example, unpredicted subvocalization might be interpreted as someone else's voice; Allen et al., 2007; Moseley et al., 2013), as well as the findings of reduced mismatch negativity in scalp-recorded electroencephalogram: i.e., reduced difference in event-related potential amplitudes between frequent (expected) and oddball (unexpected) stimuli (Wacongne, 2016). Furthermore, weakened priors could explain reduced susceptibility to some perceptual illusions (Dakin et al., 2005a; Notredame et al., 2014; King et al., 2017). Precision imbalance in favour of sensory evidence would also result in overly high learning rate and could explain the 'jumping to conclusions' observed in probabilistic reasoning tasks (Huq et al., 1988; Fine et al., 2007). Being bombarded with overly salient sensory experiences, the drive for minimization of prediction error might lead to forming compensatorily strong and idiosyncratic high-level priors (beliefs) which would explain the emergence and persistence of delusions (Fletcher and Frith, 2009; Adams et al., 2013; Corlett and Fletcher, 2014). These high-level priors might also provide another mechanism via which delusion-related hallucinations can arise: by overriding sensory evidence (Powers et al., 2016; Corlett et al., 2019).

In ASD, increased weighing of sensory information could explain reduced sensitivity to context (Mitchell and Ropar, 2004), and would subsume previous ASD theories emphasizing detail-oriented perceptual and cognitive processing style (*Weak Central Coherence:* Frith, 1989; Happé and Frith, 2006; *Enhanced Perceptual Functioning:* Mottron and Burack, 2001; Mottron et al., 2006). Furthermore, increased sensory precision could explain enhanced performance on tasks involving simple stimuli as well as sensory hypersensitivities and sensory overload (*Intense World Theory:* Markram and Markram, 2010). For instance, in the embedded figures test, where ASD exhibit superior performance, brain imaging studies have found increased visual cortical activation and decreased prefrontal

activation, consistent with sensory overwheighing and diminished top-down modulation interpretation (Ring et al., 1999; Lee et al., 2007; Manjaly et al., 2007). Similar to SCZ account, overwheighing of sensory information might also explain reduced susceptibility to some perceptual illusions in ASD (Happé, 1996; Chouinard et al., 2016; although see Ropar and Mitchell, 2001). In social situations, attending to sensory details at the expense of relevant social cues might contribute to social interaction difficulties - one of the most prominent symptoms of ASD (*Theory of Mind deficits:* Baron-Cohen et al., 1985; Baron-Cohen, 2000). The strong desire for routines and rigid behaviors observed in ASD might be understood as an attempt to introduce more predictability in one's environment, where the lack of contextual modulation results in more variable and overwhelming sensory experiences (Gomot and Wicker, 2012; Pellicano and Burr, 2012b). Furthermore, overly precise sensory representations or weaker priors would imply high learning rates, which could lead to forming prior expectations that are too specific and do not generalize across situations (Plaisted, 2001; Van de Cruys et al., 2014). In other words, the Bayesian accounts of ASD provide a computational framework for what many previous theories have captured conceptually. In particular, it allows to address one of the long-standing disagreements between top-down (Happé and Frith, 2006; Mitchell and Ropar, 2004) and bottom-up theories (Mottron et al., 2006; Markram and Markram, 2010), by providing a unifying framework that is well-suited for studying the interaction between bottom-up and top-down information flow. However, recent debates surrounding Bayesian accounts of ASD still show a similar divide (Pellicano and Burr, 2012b; Van de Cruys et al., 2013; Brock, 2012; Teufel et al., 2013; Friston et al., 2013; van Boxtel and Lu, 2013), with some authors proposing weaker priors (Pellicano and Burr, 2012b), others arguing for increased sensory precision (Brock, 2012; Van de Cruys et al., 2013; Lawson et al., 2014).

The experimental work inspired by the Bayesian accounts of ASD and SCZ has nevertheless produced mixed results (see studies summarized in **Table 1.2** and **Table 1.1**). In SCZ, intact priors have been reported across many visual illusions (Grzeczkowski et al., 2018; Kaliuzhna et al., 2019) and for repetition probability effects in face repetition suppression (Kovács et al., 2019). Most other studies investigated the relationship between priors and symptoms of delusions and hallucinations rather than the diagnostic category of schizophrenia. With this approach, some studies found delusions to be associated with weaker perceptual priors in bistable perception (Schmack et al., 2013, 2015), disambiguation of local features in two-tone images (Davies et al., 2018) and cross-modal (audio-visual) associative learning in random-dot kinematogram task (Stuke et al., 2019). However, other studies reported the opposite: delusions being associated with stronger perceptual priors in disambiguation of two-tone images (Teufel et al., 2015) and in sequential information-sampling task (Baker et al., 2019). Delusion-proneness in the general population has also been associated with stronger cognitive placebo-induced priors in bistable perception (Schmack et al., 2013), while in SCZ patients such priors were found to be weaker on the group level, but correlated positively with the severity of positive symptoms (Schmack et al., 2017). Hallucination symptoms have been associated with stronger priors in cross-modal (audio-visual) associative learning (Powers et al., 2017), disambiguation of global features in two-tone images (Davies et al., 2018), disambiguation of degraded speech (Alderson-Day et al., 2017) and auditory tone duration reproduction (Cassidy et al., 2018). More nuanced findings have been reported by three studies using more detailed computational models to explain SCZ belief updating in probabilistic reasoning tasks. Using a circular inference model where both priors and sensory evidence can be reverberated multiple times during a single belief update, Jardri et al. (2017) found positive SCZ symptoms to be related to over-counting of sensory evidence. Stuke et al. (2017) investigated non-linearity of belief updating and

found hallucination- and delusion-proneness in the general population to be associated with larger updates when presented with surprising/disconfirmatory information. Finally, using a Bayesian bielief-updating model with attractor-like dynamics, Adams et al. (2018) showed that SCZ patients could be characterized as having increased belief instability and response stochasticity.

In ASD, many studies have reported intact priors: in probability effects in repetition suppression (Ewbank et al., 2015), figure-ground segregation (Spanò et al., 2016), the direction-of-gaze prior (Pell et al., 2016), the light-from-above prior (Croydon et al., 2017), disambiguation of two-tone images (Van de Cruys et al., 2018), brightness illusion (Laeng et al., 2018), and Kaniza's triangle (Utzerath et al., 2019) and many many other perceptual illusions (Chouinard et al., 2016). Weaker priors have been reported in a probabilistic orientation discrimination task (Skewes et al., 2015), in repetition suppression (Ewbank et al., 2014, 2017), in auditory localization (Skewes and Gebauer, 2016), for central tendency bias in time interval reproduction (Karaminis et al., 2016), for the slow-speed prior (Powell et al., 2016), for adaptation to direction of social stimuli (Lawson et al., 2018), for action prediction (Chambon et al., 2017) and for disambiguation of two-tone images (Król and Król, 2019). A couple of studies also investigated the dynamics of prior updating. Lieder et al. (2019) studied contraction bias in auditory pitch discrimination and found that adults with ASD showed slower prior updating, with their priors being more influenced by distant sensory history. Lawson et al. (2017) studied prior updating in ASD under volatile conditions. In a cross-modal (audio-visual) associative learning task, they found adults with ASD to exhibit increased meta-volatility estimates, which led to smaller adjustments of their learning rate when transitioning from stable to volatile phases of the task. In contrast to Lieder et al. (2019), however, the average learning rates were not found to be different.

One methodological limitation that runs across both SCZ and ASD studies is that explicit generative models of the tested inferences are rarely implemented (see **Table 1.2** and **Table 1.1**). Consequently, in the case of positive findings, it often remains unclear, what underlying mechanisms are responsible for the observed differences. This is perfectly exemplified by studies that do fit detailed models and capture nuanced effects that would be hard to infer from the behavioral data alone (e.g., Lawson et al., 2017; Jardri et al., 2017; Stuke et al., 2017; Powers et al., 2017; Adams et al., 2018; Baker et al., 2019). In particular, very few studies have used experimental designs and computational models that can reliably quantify individual prior and likelihood effects (see **Table 1.2** and **Table 1.1**). Thus, when reduced biases are observed, they are often reported as being a result of weaker priors, even though reduced biases could also result from more precise sensory representations. Importantly, this fails to address the unresolved debate of whether Bayesian impairments in ASD originate from weaker priors or stronger likelihoods (Pellicano and Burr, 2012b; Brock, 2012; Van de Cruys et al., 2013; van Boxtel and Lu, 2013). Furthermore, while ASD and SCZ share a lot of similarities within the Bayesian framework (with the proposed primary impairment in both disorders being weaker influence of priors, relative to sensory evidence), they tend to be tested on different tasks. This leaves a gap between the theory and the empirical observations; studying these disorders together might help clarify where they computationally diverge. Finally, many studies are based on schizotypy and autistic traits in the general population, under the assumption that the proposed impairments would generalize along these dimensions. However, most studies that do test both the traits and the clinical groups find divergent effects. For instance, Ewbank et al. (2014, 2017) found repetition suppression along autistic traits in the general population to be reduced for faces, objects and scenes, while in ASD group it was reduced only for faces. Simi-

larly, Lawson et al. (2018) found reduced adaptation to direction for social and non-social stimuli along autistic traits in the general population, while in ASD reduced adaptation was found only for social stimuli. In SCZ context, Schmack et al. (2013, 2017) found placebo-induced priors along delusion-proneness in the general population to be stronger, while in a clinical SCZ sample placebo-induced priors were found to be weaker. Together, such discrepancies suggest that the assumption of continuity of the effects should not be made lightly.

Table 1.1: Recent experimental studies testing Bayesian accounts of schizophrenia (SCZ). Note that 'Type of prior' column provides descriptive, not taxonomic information. N/A - not applicable. CTR - controls. Exp 1/2/3 - denote different experiments/datasets when there is more than one.

SCZ Study	Reported effects	Computational Modelling	Disentangling priors from likelihoods	Type of prior	Sample
Stabilizing of 3D bistable rotating sphere (Schmack et al., 2013; Schmack et al., 2015; Schmack et al., 2017)	Weaker perceptual priors, in SCZ and delusion-proneness; stronger cognitive priors in delusion-proneness, weaker in SCZ	None, but fMRI used	N/A	Natural priors and placebo-induced priors	Exp 1: N = 105, delusion-proneness Exp 2: 28 SCZ, 32 CTR Exp 3: 21 SCZ, 28 CTR
Disambiguation of two-tone images (Teufel et al., 2015)	Stronger perceptual priors	None	N/A	Single-presentation visual prior	Exp 1: 18 at-risk, 16 CTR Exp 2: N = 40, delusion-proneness
Pavlovian conditioning-induced auditory hallucinations (Powers et al., 2017)	Stronger associative priors	Hierarchical Gaussian Filter	No	Frequency of cue-stimulus co-occurrence	30 Hallucinators 29 CTR
Qualitative probabilistic reasoning (Jardri et al., 2017)	Over-counting likelihood in positive symptoms; over-counting prior in negative symptoms	Circular Inference Model	Possibly, but not demonstrated	Explicitly presented on each trial via visual size representation	25 CTR, 25 SCZ
Probabilistic reasoning task with cross-modal priors (Stuke et al., 2017)	Low resilience against surprising information in delusion-proneness and in hallucination-proneness	Custom non-linear belief-updating model	No	Frequency of cue-stimulus co-occurrence	N = 94 Hallucination-proneness and delusion-proneness
Processing of ambiguous speech (Alderson-Day et al., 2017)	Stronger speech priors in non-clinical voice-hearers	None, but fMRI used	N/A	Natural priors	12 non-clinical voice-hearers, 17 CTR
Auditory tone duration reproduction task (Cassidy et al., 2018)	Stronger priors linked to hallucinations, linked to dopamine	Bayesian model (simulations only)	Possibly, but not demonstrated	Distribution of recent sensory history	Exp 1: N = 30 (CTR; pharmacological manipulation) Exp 2: 16 SCZ, 17 CTR
Disambiguation of two-tone images: local/global features (Davies et al., 2018)	Global features: stronger priors for hallucination-proneness; Local features: weaker priors for delusion-proneness	Drift-Diffusion Model and Signal Detection Theory	N/A	Single-presentation visual prior	N = 40 Hallucination-proneness and delusion-proneness
A range of classical visual illusions (Grzeczkowski et al., 2018)	Intact priors	None	N/A	Natural priors	59 SCZ, 54 CTR,
Probability estimation in a sequential-sampling task (Adams et al., 2018)	Increased belief instability and response stochasticity	Hierarchial Gaussian Filter with attractor-like dynamics	N/A	Posterior after previous sample	Exp 2: 56 SCZ, 111/60 CTR
Pavlovian conditioning-induced priors in random-dot kinematogram and in probabilistic reasoning (Stuke et al., 2019)	Weaker perceptual priors, intact cognitive priors	Logistic regression	N/A	Frequency of cue-stimulus co-occurrence	N = 123, delusion-proneness
Perception of brightness, line length, and motion direction (Kaliuzhna et al., 2019)	Intact priors	None	N/A	Natural priors	19 SCZ, 21 CTR.
Incentivized information-sampling task (Baker et al., 2019)	Stronger priors in high-delusion patients	Partially observable Markov decision process	Possibly, but not demonstrated	Posterior after previous sample	24 SCZ, 21 CTR
Repetition probability effects (Kovács et al., 2019)	Intact priors	None	N/A	Frequency of repetition/alteration	17 SCZ, 17 CTR

Table 1.2: Recent experimental studies testing Bayesian accounts of ASD. Note that 'Type of prior' column provides descriptive, not taxonomic information. N/A - not applicable, AQ - Autism Quotient, CTR - controls. Exp 1/2/3 - denote different experiments/datasets when there is more than one.

ASD Study	Reported effects	Computational Modelling	Disentangling priors from likelihoods	Type of prior	Sample
Repetition suppression (Ewbank et al., 2014, 2017)	Weaker priors for faces in ASD; Weaker for faces, shapes and scenes in AQ	None, but fMRI used	N/A	Priming/adaptation	AQ traits N = 27/29/31 (3 experiments) 15 ASD, 15 CTR (adults)
Adaptation to numerosity (Turi et al., 2015)	Weaker priors	None	N/A	Adaptation, presented just before the test trial	16 ASD, 18 CTR (children)
Orientation discrimination (Skewes et al., 2015)	Weaker priors	Signal Detection Theory	N/A	Average outcome frequency	AQ traits N = 29
The effect of expectations on repetition suppression of faces (Ewbank et al., 2015)	Intact priors	None, but fMRI used	N/A	Frequency of stimulus presentation	AQ traits N = 32
Figure-ground segregation (Spanò et al., 2016)	Intact priors	None	N/A	Natural priors (low-level: convexity; high-level: familiar objects)	23 ASD, 30 CTR (children and adults)
Direction of gaze (Pell et al., 2016)	Intact priors	None	N/A	Natural prior	AQ traits N = 34
Slow-speed prior (Powell et al., 2016)	Weaker priors	None (but uses math. derivations)	N/A	Natural prior	AQ traits N = 31 (Exp 1) N = 26 (Exp 2)
Susceptibility to simple optical illusions (Chouinard et al., 2016)	Weaker priors only for two illusions (orientation of simple shapes)	None	N/A	Natural prior	AQ traits N = 131
Central tendency bias in time interval reproduction (Karaminis et al., 2016)	Weaker priors	Bayesian model (simulations only)	Possibly, but not demonstrated	Distribution of recent sensory history	78 CTR (children and adults) 23 ASD (children)
Auditory localization (Skewes and Gebauer, 2016)	Weaker priors	Signal Detection Theory	N/A	Average frequency of different stimuli	16 ASD, 19 CTR (adults)
Adaptation to audiovisual asynchrony (Turi et al., 2016)	Weaker priors	None	N/A	Adaptation/recalibration	16 ASD, 16 CTR (adults)
Social priors during action prediction (Chambon et al., 2017)	Weaker priors	Signal Detection Theory	N/A	Frequency of intention-action contingencies	18 ASD, 20 CTR (adults)
Light-from-above (Croydon et al., 2017)	Intact priors	None	N/A	Natural prior	18 ASD, 18 CTR (children)
Volatile audio-visual Pavlovian conditioning (Lawson et al., 2017)	Higher meta-volatility	Hierarchical Gaussian Filter	No	Frequency of cue-stimulus co-occurrence	24 ASD, 25 CTR (adults)
Disambiguation of two-tone images (Van de Cruys et al., 2018)	Intact priors	None	N/A	Single-presentation visual prior	AQ traits N = 282
Pupillary responses to illusions of brightness (Laeng et al., 2018)	Intact priors	None	N/A	Natural prior	11 ASD, 24 CTR (adults)
The effect of expectations on repetition suppression of objects (Utzerath et al., 2018)	Weaker priors	None	N/A	Adaptation; frequency of stimulus presentation	22 ASD, 22 CTR (adolescents)
Adaptation to direction for social and non-social stimuli (Lawson et al., 2018)	Weaker priors only for social stimulus in ASD and for all stimulus in AQ	None	N/A	Adaptation/recalibration	AQ traits N = 28 (Exp 1) 17 ASD, 19 CTR (adults; Exp 2)
4 tasks: Illusory contours, blur detection, Mooney, representational momentum (Tulver et al., 2019)	Intact priors	None	N/A	Different type of prior in each task	AQ traits N = 44
Kanisza's triangle – illusory contours (Utzerath et al., 2019)	Intact priors	None, but fMRI used	N/A	Natural prior	22 ASD, 22 CTR (adolescents)
Contraction bias in tone discrimination (Lieder et al., 2019)	Slow updating, slow forgetting	Psychometric curves	N/A	Mean of a feature (recent sensory history)	Exp 2: 37 ASD, 32 CTR Exp 3: 16 ASD, 26 CTR (adults)
Disambiguation of two-tone images (Król and Król, 2019)	Weaker priors	Eye-tracking only	N/A	Single-presentation visual prior	23 ASD, 21 CTR (adolescents and adults)
Cue-action association effects on action prediction (Amoruso et al., 2019)	Weaker priors	Signal Detection Theory	N/A	Cue-action contingencies	24 ASD, 24 CTR (children)
Low-level visual spacial context modulation (Sandhu et al., 2020)	Intact priors	Psychometric curves	N/A	Natural priors	27 ADS, 20 CTR (adults)

1.4 Aims and outline of the thesis

The aim of this thesis was to experimentally investigate the Bayesian accounts of SCZ and ASD by addressing some of the methodological limitations of the previous work. To this end, we studied SCZ and ASD patient groups as well as schizotypy and autistic traits in the general population within a single visual motion estimation task that induces perceptual priors (Chalk et al., 2010). An important important feature of this task design was that responses were on an interval scale (unlike most previous work in this area that used binary response design), which allowed to estimate perceptual variability, which in turn made it possible to disentangle prior and likelihood influences on perception. Another important feature of the used task design was that the acquired prior was expressed in multiple behavioral measures (e.g., detection, discrimination, reaction times). This allowed to detect more nuanced differences in how the acquired priors influenced performance in the task and gave better control over potential confounds. Finally, to stay true to the principles of computational psychiatry, we performed extensive computational modelling analysis, which allowed to quantify nuanced individual and group differences in terms of model parameters.

The first part of the thesis (**Chapters 2-4**) consists mostly of published work and constitutes the core part of the thesis. Each chapter is based on the same experimental design (using the Moving Dots task described above together with Bayesian modelling) but in each case it is applied to a novel dataset: **Chapter 2** - autistic and schizotypy traits in the general population, **Chapter 3** - SCZ patient group, **Chapter 4** - ASD patient group. The second part of the thesis (**Chapters 5** and **6**) presents further exploratory analysis focused on additional effects observed across the three aforementioned datasets. This serves to control for potential confounds, allows to investigate other individual and group differences and relates our work to other frameworks and areas of research. In **Chapter**

5 we re-analyse the performance in the task by accounting for additional biases present in our data that the standard Bayesian model could not account for. This relates our work to research on *reference repulsion*. In **Chapter 6** we investigate the observed Bayesian effects by accounting for within-trial dynamics. This leads us to constructing a Continuous Choice Drift Diffusion Model (CDM) - an extension of the classical binary choice Drift Diffusion Model (DDM).

Part I

Perceptual Bayesian inference in ASD and SCZ

2

Perceptual Bayesian inference in autistic and schizotypy traits

This chapter includes a postprint of a published journal article: 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.

Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration

Povilas Karvelis[1], Aaron R. Seitz[2], Stephen M. Lawrie[3,4] and Peggy Seriès[1,*]

1- IANC, School of Informatics, University of Edinburgh, Edinburgh, UK

2- UC Riverside, Department of Psychology, Riverside, CA, USA

3- Division of Psychiatry, University of Edinburgh, Edinburgh, UK

4- Patrick Wild Centre, University of Edinburgh, Edinburgh, UK

*** Corresponding author: pseries@inf.ed.ac.uk**

Abstract

Recent theories propose that schizophrenia/schizotypy and autistic spectrum disorder are related to impairments in Bayesian inference i.e. how the brain integrates sensory information (likelihoods) with prior knowledge. However existing accounts fail to clarify: i) how proposed theories differ in accounts of ASD vs. schizophrenia and ii) whether the impairments result from weaker priors or enhanced likelihoods. Here, we directly address these issues by characterizing how 91 healthy participants, scored for autistic and schizotypal traits, implicitly learned and combined priors with sensory information. This was accomplished through a visual statistical learning paradigm designed to quantitatively assess variations in individuals' likelihoods and priors. The acquisition of the priors was found to be intact along both traits spectra. However, autistic traits were associated with more veridical perception and weaker influence of expectations. Bayesian modeling revealed that this was due, not to weaker prior expectations, but to more precise sensory representations.

Introduction

In recent years Bayesian inference has come to be regarded as a general principle of brain function that underlies not only perception and motor execution, but hierarchically extends all the way to higher cognitive phenomena, such as belief formation and social cognition. Impairments of Bayesian inference have been proposed to underlie deficits observed in mental illness, particularly schizophrenia [1-3, 49-51] and autistic spectrum disorder (ASD) [4-7]. The general hypothesis for both disorders is that the weight, also called "precision", ascribed to sensory evidence and prior expectations is imbalanced, resulting in sensory evidence having relatively too much influence on perception.

In schizophrenia, overweighting of sensory information could explain the decreased susceptibility to perceptual illusions[8], as well as the peculiar tendency to jump to conclusions[9]. Moreover, the systematically weakened low-level prior expectations might lead to forming compensatory strong and idiosyncratic high-level priors (beliefs), which would explain the emergence and persistence of delusions as well as reoccurring hallucinations [1-3].

In ASD, the relatively stronger influence of sensory information could explain hypersensitivity to sensory stimuli and extreme attention to details. The weaker influence of prior expectations would also result in more variability in sensory experiences. The desire for sameness and rigid behaviors could then be understood as an attempt to introduce more predictability in one's environment [4]. Furthermore, this could lead to prior expectations which are too specific and which do not generalize across situations [5]. While all theories agree that the relative influence of prior expectations is weaker in ASD, the primary source of this imbalance is debated: does it arise from increased sensory precision (i.e. sharper likelihood) or from reduced precision of prior expectations? [10-12] (**Fig. 1**). Some authors argue for attenuated priors [4, 11], while others argue for increased sensory precision [6, 7, 10, 13] but conclusive experimental evidence is lacking.

A number of studies have aimed at testing Bayesian theories, either in a clinical population, or by studying individual differences in the general population [14-17] under the hypothesis of a continuum between autistic/schizotypal traits and ASD/schizophrenia [18-20]. Attenuated slow-speed priors were reported in a motion perception task in individuals with ASD traits [14]. Autistic

children also showed attenuated central tendency prior in temporal interval reproduction[21]. Attenuated priors were also reported in perceptual tasks that incorporate probabilistic reasoning [15, 22]. However, the direction of gaze priors [23] and the light-from-above priors [24] were found to be intact. Autistic children also demonstrated intact ability to update their priors in a volatile environment in a decision-making task [25] but a follow-up study in ASD adults showed that they overestimate volatility in a changing environment [26]. In schizophrenia/schizotypal traits, Teufel et al.[16] reported increased influence of prior expectations when disambiguating two-tone images, while Schmack et al.[27,28] reported weakened influence of stabilizing predictions when observing a bistable rotating sphere.

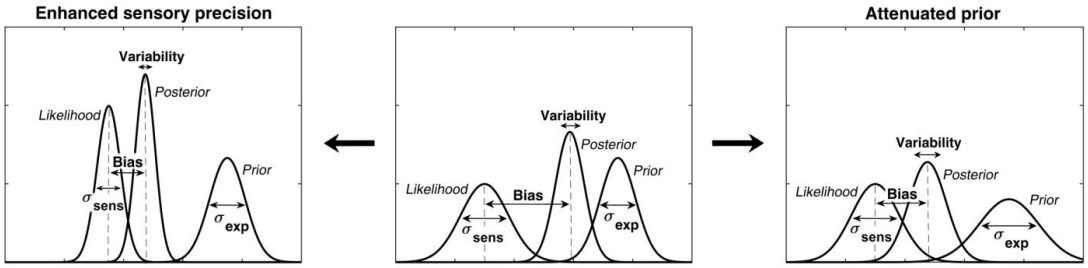


Figure 1. Alternative hypotheses for ASD impairments within the Bayesian inference framework. In Bayesian terms, the percept can be described as a posterior distribution, which is a combination of sensory information (likelihood) and prior expectations (prior). Two contrasting hypotheses have been proposed to underlie behavioral differences in ASD: enhanced sensory precision, i.e. smaller σ_{sens} (left) vs. attenuated priors, i.e. larger σ_{exp} (right). Both hypotheses predict a reduced influence (bias) of the prior on the location of the posterior distribution (posterior mean). However, these alternatives differ in their predictions for perceptual variability, which is determined by the posterior width: the enhanced sensory precision hypothesis should lead to reduced variability while the attenuated prior hypothesis should lead to increased variability. By measuring both bias and variability, our experimental paradigm can distinguish between these two hypotheses.

Overall, the existing findings are not only mixed, but also employ very different paradigms, which makes their direct comparison difficult. Further, a critical limitation of most studies (except for Karaminis et al. [21]) is the lack of formal computational models that can test whether behavioral differences originate from different priors or from different likelihoods. Moreover, to our knowledge, despite the similarity of the Bayesian theories proposed for ASD and schizophrenia, there is no previous work investigating both autistic and schizotypal traits within the same experimental paradigm so as to test their differences.

We here address these questions empirically in a context of visual motion perception. We used a previously developed statistical learning task[29] in which participants have to estimate the direction of motion of coherently moving clouds of dots (**Fig. 2**). Chalk et al. [29] found that in this task healthy participants rapidly and implicitly develop prior expectations for the most frequently presented motion directions. This in turn alters their perception of motion on low contrast trials resulting in attractive estimation biases towards the most frequent directions. In addition, prior expectations lead to reduced estimation variability and reaction times, as well as increased detection performance for the most frequently presented directions. When no stimulus is presented, the acquired expectations sometimes lead to false alarms (hallucinations), again, mostly in the most frequent directions. Importantly, such biases were well described using a Bayesian model, where participants acquired a perceptual prior for the visual stimulus that is combined with sensory information and influences their perception. As such, this paradigm is well suited to quantitatively model variations in likelihoods and priors in individuals with ASD or schizotypal traits.

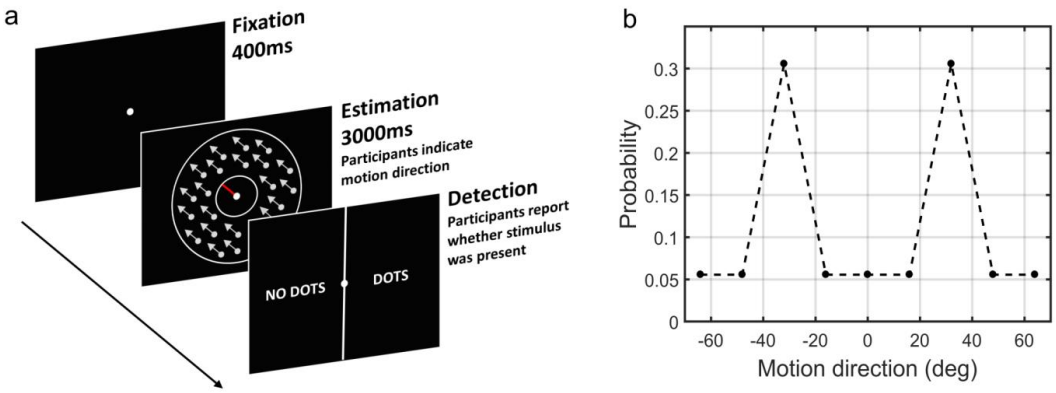


Figure 2. The moving dots task. (**a**) Sequence of events on a single trial. First, a fixation point is presented. Next, a field of coherently moving dots is presented along with an estimation bar (extending from the fixation point) which participants are required to move to indicate perceived motion direction. Lastly, in a two-alternative forced choice, participants are asked to report whether they saw the dots during the estimation part (detection task). (**b**) The probability of different motion directions being presented: directions at ±32° are presented more often than other directions. Motion direction is plotted relative to a central reference angle (at 0°), which was randomly set for each participant.

Results

Here, we investigated individual differences in statistical learning in relation to autistic and schizotypal traits in a sample of 91 healthy participants. 8 participants failed to perform the task satisfactorily and were excluded from the analysis (see *Methods*), leaving 83 participants in the study (41 women and 42 men, age range: 18-69; mean: 25.7).

Task behavior at low contrast

First, we investigated whether participants acquired priors on the group level. We discarded the first 170 trials as that is how long it took for the 2/1 and 4/1 staircases contrast levels to converge (**Appendix 1—Figure 2**) and for prior effects to become significant (**Appendix 1—Figures 3, 4 and 5**). We analyzed task performance at low contrast levels (converged 2/1 and 4/1 staircases contrast levels) where sensory uncertainty is high. Replicating findings of Chalk et al. (2010), we found that on the group level people acquired priors that approximated the statistics of the task. Such priors were indicated by: attractive biases towards ±32° (**Fig. 3a**), less variability in estimations at ±32° (**Fig. 3b**; standard deviation of estimations 11.9± 0.30° at ±32° versus 13.84±2.38° over all other motion directions; signed rank test: $p < 0.001$), shorter estimation reaction times at ±32° as compared to all other motion directions (**Fig. 3c**; average reaction time was 201.87 ± 2.47 ms at ±32° versus 207.75 ± 2.60 ms over all other motion directions; signed rank test: $p < 0.001$) and better detection at ±32° as compared to all other motion directions (**Fig. 3d**; detected 75.57 ± 0.65% at ±32° versus 66.70 ± 0.83% over all other motion directions; signed rank test: $p < 0.001$).

No-stimulus performance

Another indicator of acquired priors is the distribution of estimation responses on trials when no actual stimulus was presented. We found that participants sometimes still reported seeing dots (experienced hallucinations) but mostly so around ±32° (**Fig. 3f**, solid line). To quantify the statistical significance of hallucinations around ±32°, the space of possible motion directions was divided into 23 bins of 16° and the probability of estimation within 8° of ±32° was multiplied by

the total number of bins:

$$\text{Prel} = p(\theta_{est} = \pm 32(\pm 8)°) \cdot N_{bins}\,, \qquad (1)$$

where N_{bins} is the number of bins (45), each of size 16°. This probability ratio would be equal to 1 if participants were equally likely to estimate within 8° of ±32°, as they were to estimate within other bins. We found that the median of Prel was significantly greater than 1 (median(Prel) = 1.6, $p<0.001$, signed rank test). Furthermore, the estimation distribution when no dots where detected (**Fig. 3f**, dash-dot line) was found to be significantly flatter (median(Prel) = 0, $p < 0.001$, signed rank test comparing with the median of Prel for hallucinations), suggesting that the hallucinations were indeed of perceptual nature (rather than related to a response bias).

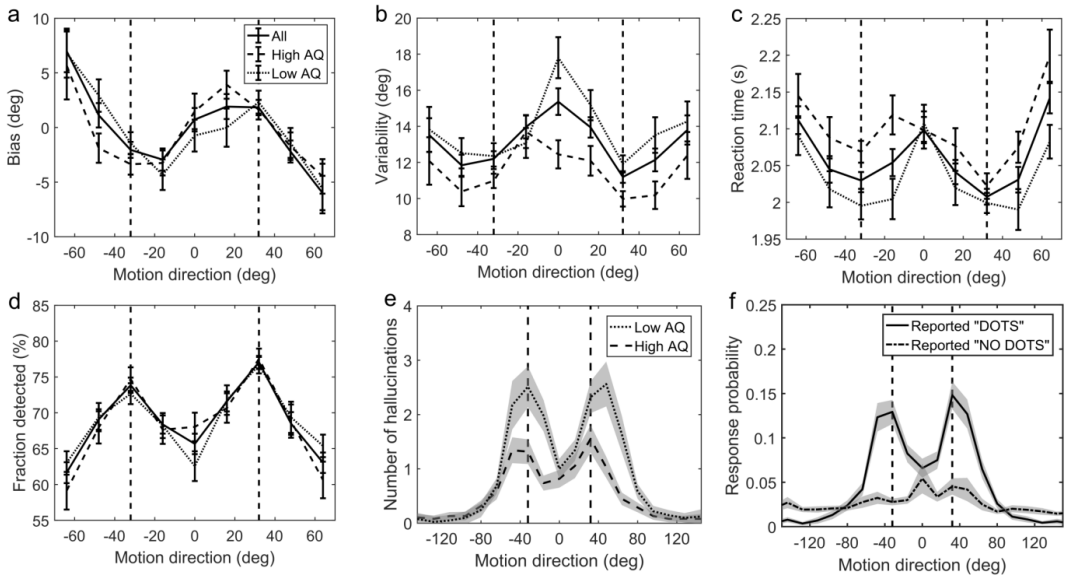


Figure 3. Average group performance on low-contrast trials (**a-d**) and on trials with no stimulus (**e**). (**a**) Mean estimation bias, (**b**) standard deviation of estimations, (**c**) estimation reaction time and (**d**) fraction of trials in which the stimulus was detected. (**f**) Probability distribution of estimation responses on trials without stimulus. The solid line denotes the estimation responses when participants reported detecting a stimulus (hallucinations). The dash-dot line denotes estimation distributions when participants correctly reported not detecting a stimulus. (**e**) Distribution of hallucinations for high and low AQ groups (median split). The vertical dashed lines correspond to the two most frequently presented motion directions (±32°). Error bars and shaded areas represent within-subject standard error.

Task performance and autistic/schizotypy traits

Participants were prescreened to make sure they covered a wide range of autistic and schizotypy scores. The AQ scores in our sample ranged from 6 to 41 (full questionnaire range being 0 to 50) with a mean (±SD) of 20.3 (±8.3), which is a typical distribution for a neurotypical population[57]. The RISC scores ranged from 8 to 55 (full questionnaire range being 0 to 78) with a mean of 31.7 (±11.9), and the SPQ scores ranged from 4 to 59 (full questionnaire range being 0 to 74) with a mean of 26.4 (±13.8), both of which are typical for the general population[41,42].

We found that on low contrast trials autistic traits lead to less variability in estimations (**Fig. 4b**; mean standard deviation of estimations: $r = -0.327$, $p < 0.001$), which remained significant after Bonferroni correction (p = 0.002). Moreover, there was a negative relationship between autistic traits and estimation bias, which was trending according to robust regression (**Fig. 4a**; mean absolute estimation bias: $r = -0.175$, p = 0.053) and significant according to Kendall's correlation ($\tau_b = -0.163$, p = 0.032), however, it did not survive Bonferroni correction (p = 0.212). In the Bayesian framework, less bias could arise either due to wider priors or narrower sensory likelihoods, while less variability could be a result of either narrower priors or narrower likelihoods (see **Fig. 1**). Thus, observing less bias and less variability together suggests that the effects are driven by narrower likelihoods. An alternative is that the differences in variability could be due to differences in motor precision, which we further assess via modeling (below).

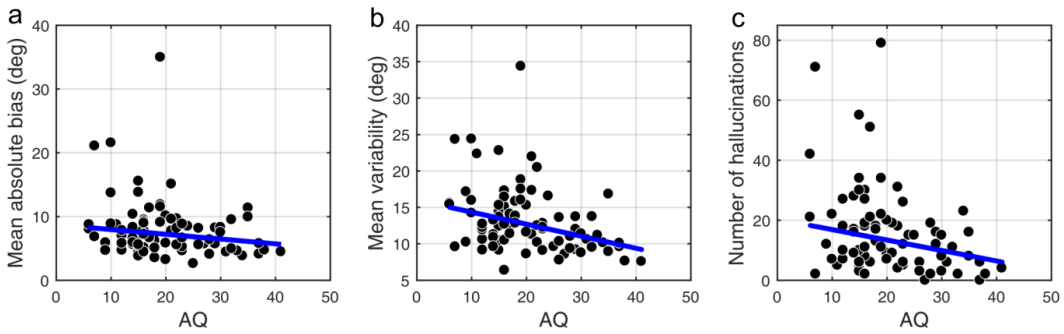


Figure 4. Correlations between AQ scores and task performance on low contrast trials (**a, b**) and when no stimulus is presented (**c**). (**a**) Mean absolute bias (r = −0.175, p = 0.053), (**b**) mean standard deviation (i.e. variability) of estimations (r = −0.327, p < 0.001), and (**c**) the total number of hallucinations (r = −0.238, p = 0.010). The blue lines are robust regression slopes.

Schizotypy traits (RISC and SPQ scores) did not show any effect on task performance at low contrast as indicated by the absence of correlations with mean absolute estimation bias (RISC: r = 0.140, p = 0.197; SPQ (N=39): r = −0.160, p = 0.204) and with mean estimation variability (RISC: r = 0.197, p = 0.092; SPQ (N=39): r = −0.229, p = 0.171); see **Appendix 1—Figures 6, 7 and 8**.

No-stimulus trials and autistic/schizotypal traits

We also investigated how the traits affected performance on trials when no actual stimulus was presented. First, we looked at the total number of estimations. We found that autistic traits were associated with less hallucinations (**Fig. 4c**; r = −0.238, p = 0.010), while schizotypal traits were found to have no effect on the number of hallucinations (RISC: r = 0.126, p = 0.163; SPQ (N=39): r = -0.010, p = 0.959). Secondly, we looked for relationships between the traits and how the estimations on no-stimulus trials were distributed. Specifically, we were interested in whether the traits predicted how densely hallucinations were distributed around ±32°, as this could be considered to reflect the differences in the width of the underlying acquired prior distribution. For weaker priors we would expect a more spread out distribution of hallucinations. To test this hypothesis, we looked at the fraction of total hallucinations in the region around ±32° for three different-sized windows: Within 8°, within 16° and within 24° of ±32°. Bayesian Kendall correlation analysis on these measures provided positive evidence that none of the traits had any effect on how hallucinations were distributed, suggesting no differences in the acquired prior distributions (fraction of hallucinations within 8° of ±32°: AQ - τ_b = 0.003, BF_{01} = 7.24; RISC - τ_b = -0.050, BF_{01} = 3.73; SPQ - τ_b = 0.101, BF_{01} = 8.72; within 16° of ±32°: AQ - τ_b = -0.068, BF_{01} = 2.86; RISC - τ_b = -0.129, BF_{01} = 0.84; SPQ - τ_b = 0.018, BF_{01} = 5.45; within 24° of ±32°: AQ - τ_b = 0.057, BF_{01} = 11.67; RISC - τ_b = -0.078, BF_{01} = 2.40; SPQ - τ_b = 0.006, BF_{01} = 5.02).

Modeling results

Group level results

To quantitatively evaluate the relationships between underlying perceptual mechanisms and task performance we fitted a range of generative models. One class of models was Bayesian - it was based on the assumption that participants combine prior expectations with uncertain sensory information on a single trial basis (**Fig. 5**).

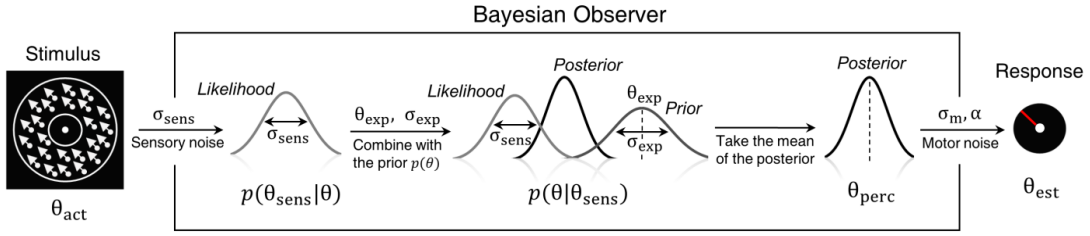


Figure 5. Bayesian model of estimation response for a single trial. The actual motion direction (θ_{act}) is corrupted by sensory uncertainty (σ_{sens}), and then combined with prior expectations (mean θ_{exp} and uncertainty σ_{exp}) to form a posterior distribution. The perceptual estimate (θ_{perc}) is defined as the mean of the posterior distribution. Finally, motor precision ($1/\sigma_m^2$) and a probability of random response (α) are incorporated to generate the response (θ_{est}). This results in 4 free model parameters: σ_{sens}, σ_{exp}, θ_{exp} and α. The motor precision is estimated from high contrast trials and is used as a fixed parameter.

To account for the possibility that the bimodal probability distribution of the stimuli, in addition to inducing prior expectations, has also affected the sensory likelihood, we constructed three variations of the Bayesian model: 'BAYES', where the sensory precision was constrained to be the same across all presented motion directions, 'BAYES_varmin', where the sensory precision was allowed to be different for the most frequently presented motion directions, but was the same across all other directions, and 'BAYES_var', where sensory precision was allowed to be different across all motion directions. Another class of models was based on the assumption that task performance can be explained by response strategies that do not involve Bayesian inference. That is, on any given trial participants responded based on the prior expectations or sensory information alone. We considered four variations of response strategy models: 'ADD1', 'ADD2', 'ADD1_m' and 'ADD2_m' (see Methods for details).

To compare the models, we computed BIC values for each individual for each model; we used individual BIC values as a summary statistic and compared the models using signed rank test in order to preserve individual variability, which corresponds to a random effects Bayesian model selection procedure. We found that the BAYES model had significantly smaller BIC values than the remaining models (see the p-values within **Fig. 6a**).

To determine how the best fitting model compared to the actual data, we analyzed the estimation biases and variation in estimation responses as predicted by BAYES (**Fig. 6b,c**). As in the experimental data analysis, we computed estimation distributions predicted by the model by assuming occasional random estimations (see Eq. (2)). Finally, using the BAYES model, we reconstructed the priors acquired by participants. While on the individual level there was a considerable variation in the shape of acquired priors (see **Appendix 1—Figure 10**), on the group level, it approximated the statistics of the task (**Fig. 6d**).

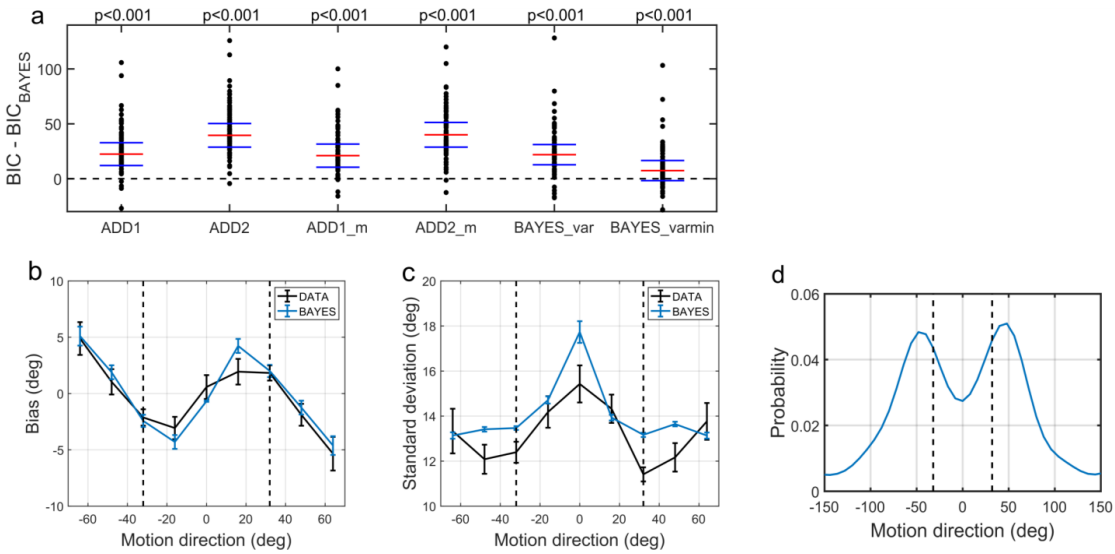


Figure 6. Modelling results. (**a**) Model comparison for all participants using Bayesian Information Criterion (BIC). y-axis measures the relative difference between BIC of each model (as indicated on the x-axis) and BIC of BAYES model. Values greater than zero on the y-axis indicate that the BAYES model provided a better fit. Each dot represents a participant. Red horizontal lines denote median values; blue horizontal lines denote 25th and 75th percentiles. p-values above the plot indicate whether the median of the difference was significantly different from zero for each model (signed rank test). Panels (**a**) and (**c**) present task performance at different motion directions as predicted by BAYES model: (**b**) estimation bias, (**c**) standard deviation of estimations. Error bars represent within-subject standard error. (**d**) Population averaged prior as recovered via BAYES model. The vertical dashed lines correspond to the two most frequently presented motion directions (±32°).

Model parameters and autistic/schizotypal traits

Correlational analysis of BAYES model parameters showed that there was no correlation between AQ and the precision of the prior σexp (**Fig. 7b**; r = 0.018, p = 0.962). That autistic traits had no effect on the precision of the prior was confirmed by Bayesian Kendall correlation, which provided positive evidence (τ_b = 0.001, BF_{01} = 6.99).

Importantly, autistic traits were found to be strongly associated with less uncertainty in the sensory likelihood, σsens (**Fig. 7c**; r = −0.185, p = 0.011), which also remained significant after Bonferroni correction (p = 0.044). Finally, there was no correlation with the amount of random estimations (**Fig. 7d**; r = −0.135, p = 0.238). Motor precision, which was estimated from high contrast trials, separately from all other parameters (see Methods), was also correlated with autistic traits (r = 0.245, p = 0.012). On the other hand, consistent with the absence of differences in the behavioral findings, schizotypal traits were not associated with any difference in the BAYES model parameter values **(Appendix 1—Figure 9)**, and in particular, were found to have no effect on prior precision (RISC: τ_b = -0.012, BF_{01} = 6.90; SPQ: τ_b = 0.071, BF_{01} = 3.97).

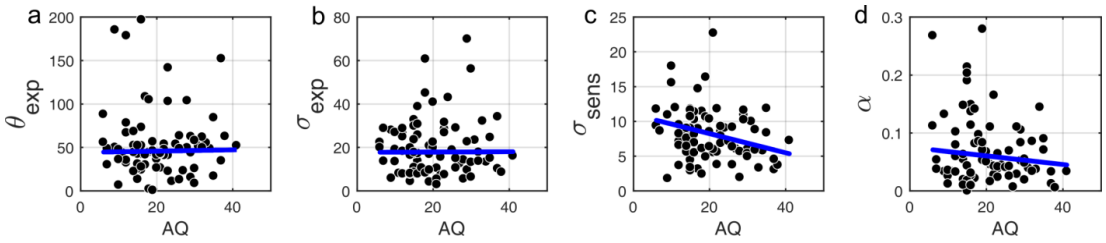


Figure 7. Correlations between AQ scores and BAYES model parameters. (**a**) θ_{exp} - mean of the prior expectations (r = 0.031, p = 0.820), (**b**) σ_{exp} - uncertainty of the prior distribution (r = 0.018, p = 0.962), (**c**) σ_{sens} - uncertainty in the sensory likelihood (r = −0.185, p = 0.011) and (**d**) α - fraction of random estimations (r = −0.135, p = 0.238). The blue lines are robust regression slopes.

Parameter recovery for BAYES

Finally, to further investigate that in our experimental paradigm the influence of stronger likelihoods can be distinguished from that of weaker priors [10, 11] we performed parameter recovery for the winning BAYES model. Parameter recovery involves generating synthetic data with

different sets of parameters (’actual parameters’) and then fitting the same model to estimate the parameters (’recovered parameters’) that are most likely to have produced the data. If actual and recovered parameters are in a good agreement, it means that the effects of different parameters can be reliably distinguished. At the same time, parameter recovery is also affected by the parameter estimation methods and even more so by the amount of data used for model fitting. Therefore, parameter recovery provides an overall check for the reliability of modelling results and is recommended as an essential step in computational modelling approaches [30].

We found that overall BAYES model (and MLE parameter estimation using simplex optimization function) recovered parameters very well, which was reflected in Pearson’s correlation between actual and recovered estimates being r > 0.9 for all model parameters (**Fig. 8**).

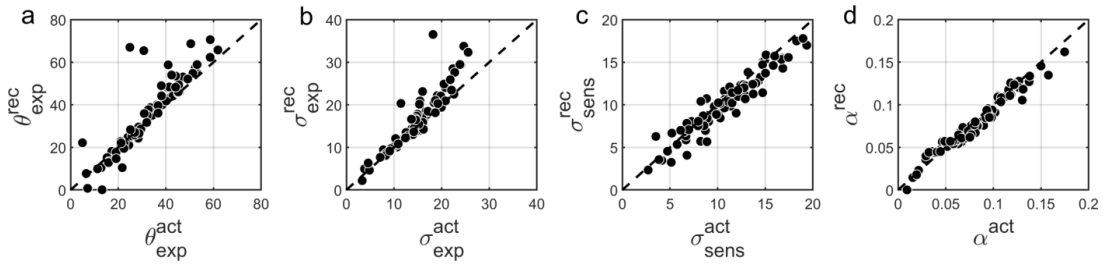


Figure 8. Comparison of actual (x-axis) vs. recovered (y-axis) parameters using the ’BAYES’ model. (**a**) θ_{exp} - mean of the prior expectations (r = 0.90), (**b**) σ_{exp} - uncertainty of the prior distribution (*r* = 0.92), (**c**) σ_{sens} - uncertainty in the sensory likelihood (r = 0.95), (**d**) α - fraction of random estimations (r = 0.98). The dashed diagonal line is a reference line indicating perfect parameter recovery.

Discussion

In this study, we investigated whether autistic and schizotypal traits are associated with differences in the implicit Bayesian inference performed by the brain. Specifically, we wanted to know whether autistic and schizotypal traits are accompanied by 1) differences in how the priors are updated and/or in their precision and/or by 2) differences in the precision with which the sensory information (the likelihood) is represented. We used a visual motion estimation task [29] that induces implicit prior expectations via more frequent exposure of two motion directions (±32°). We found that on the group level (N=83) participants acquired prior expectations towards ±32° motion directions. This was indicated by shorter estimation reaction times and better detection at ±32°, as

well as attractive biases towards ±32° and reduced estimation variability at ±32°. Moreover, when no stimulus was presented, participants sometimes still reported seeing the stimulus, mostly around ±32°. Performance was best explained by a simple Bayesian model, which provided a good fit to the data and captured the characteristic features of perceptual bias and variability. This model provided estimates of Bayesian priors and sensory likelihoods for each participant, which were then analyzed in relation to participants' schizotypal and autistic traits.

Schizotypal traits were found to have no measurable effect on perceptual biases in our task and, therefore, were not associated with any differences in the precision ascribed to priors and likelihoods. This finding challenges recent accounts of positive symptoms of schizophrenia that predict impaired updating of priors and an imbalance in precision ascribed to sensory information and prior expectations [1-3]. An immediate explanation might be that the influence of schizotypal traits in the healthy population is not strong enough to lead to behavioral differences, even if the dimensionality assumption holds. This would need to be addressed by further research investigating clinical populations. Another possibility is that the aberrant perception subconstruct of schizotypal traits, for which we did not acquire explicit measures, is more relevant for the hypothesized effects then the entire construct as a whole. For example, a recent study by Powers et al[31] found that overweighing of perceptual priors was specifically linked to hallucinatory propensity and not to the diagnostic status of psychosis itself. Furthermore, Teufel et al.[16] also found that stronger influence of prior knowledge was primarily associated with hallucinatory propensity and not with delusional propensity (i.e. the latter had no significant effects once the former was partialed out). Another possible difference between Teufel et al.[16] study and ours might be the level at which the priors operate. In Teufel et al.[16], participants were presented with ambiguotwo-tone versions of images before and after seeing the actual images in full color and had to report whether the presented two-tone image contains a face. The low-level prior for basic perceptual features (as induced in our task) might function at a hierarchically lower level than prior knowledge related to complex collection of features and semantic content (faces). The level at which prior expectations are induced has indeed been shown to matter. A series of studies by Schmack et al.[17, 27, 28] using 3D rotating cylinders report weaker low-level (perceptually-induced - stabilizing) priors but stronger high-level (cognitively-induced) priors in both schizophrenia and schizotypal traits. It is difficult to compare and reconcile these findings with ours. One possibility is that the priors induced in our task lie in between their perceptual and cognitive levels. The taxonomy of priors in relation to their place in the computational hierarchy or to their complexity or specificity is still far from being established [32] and thus the potential relevance of such

distinctions is still not known.

Autistic traits were associated with significant behavioral differences: weaker biases and lower variability of direction estimation on low contrast trials. Modeling revealed that this was because of increased sensory precision as well as higher motor precision, while there was no attenuation of acquired priors. Parameter recovery analysis confirmed that our methodology provides reliable parameter estimates and, in particular, allows disentangling variations in priors and likelihoods.

Autistic traits were also found to be associated with less false detections (hallucinations) on trials when no stimulus was presented, consistent with the idea that prior expectations had less influence in individuals with higher AQ. In an attempt to measure those individual differences, we fitted a more sophisticated Bayesian model that could account not only for the estimation performance but also for the detection data (see **Appendix 2**). This model provided a good fit to both estimation and detection data and preserved the correlation between ASD traits and the precision of the motion direction likelihood (r = -0.202, p = 0.029). However, parameter recovery was not as good as for the BAYES model presented above (see **Appendix 2 – Figure 3**) and for this reason we focused on the simpler model in this paper.

Overall, our findings are in agreement with most of the recent Bayesian theories of ASD, namely, that autistic traits are associated with a relatively weaker influence of prior expectations. However, we find that this is due to enhanced sensory precision [6, 7, 10, 13], rather than attenuated priors per se [4]. Other empirical studies inspired by the Bayesian accounts have reported either attenuated or intact priors, but most are subject to methodological limitations, either because they did not use computational modeling [15, 22,-24] or because their model could not extract likelihoods and quantify their variations [14, 26].

The idea that sensory processing could be enhanced in autism has long been proposed outside the Bayesian framework. Autistic traits have been associated with enhanced orientation discrimination [33], but only for first-order (luminance-defined) stimulus [34]. This enhancement has been proposed to be a result of either enhanced lateral [34], or a failure to attenuate sensory signals via top-down gain control [6], both of which could be directly related to narrower likelihoods in the Bayesian framework[35]. However, in motion perception, previous research did not find improved discrimination for first-order stimulus in autism, while for second-order (texture-defined) stimulus, the autistic group was found to underperform [36]. Our findings challenge these results and call for

more research in this area.

In ASD as in schizotypy, prior integration might function differently at different levels of sensory processing. For example, Pell et al.[23] reported intact direction-of-gaze priors for healthy individuals with high autistic traits and for highly functional individuals with a clinical diagnosis. The authors did not directly investigate differences in sensory precision, but the lack of behavioral differences suggests that there was none. Arguably, their paradigm involves more complex stimuli than used in our task, which are also strongly associated with semantic content (faces). It would not be surprising if increased sensory precision does not extend to such stimuli. In fact, autistic individuals are known to exhibit differential performance based on the complexity of the stimulus [34], which also lies at the foundation of some theoretical accounts, such as the 'Weak Central Coherence' [37].

In our paradigm people acquire prior expectations very quickly, within 200 trials (see **Appendix 1**), which did not allow us to study individual differences in the rate at which the priors are acquired. Bayesian accounts predict differences in the dynamical updating of the priors, namely, that both autistic and schizotypal traits should be associated with increased learning rate - which is the ratio of likelihood and posterior precisions [7]. Our findings of increased sensory precision in autistic traits also suggest that their learning rate should be faster. However, this prediction might need to be more nuanced for volatile environments when there are multiple (hierarchical) levels of uncertainty that need to be updated simultaneously. A recent study by Lawson et al.[26] found that when transitioning from stable to volatile environments, autistic adults showed larger change in the learning rate about volatility and smaller change in the learning rate about the environmental probabilities, while the average learning rates were found to not be different from those of controls.

Another aspect that our paradigm could not test is the specificity of the acquired priors [32]. Some Bayesian accounts [5] predict that priors may be overly context-sensitive in autism. This is in line with the view that generalization is impaired in autism [38]. Furthermore, such over-specificity is thought to be stronger with more repetitive stimuli [39]. Future research could address this using statistical learning paradigms that incorporate increasingly distinct contexts or stimuli.

Conclusion

We investigated statistical learning and Bayesian inference in a visual motion perception task along autistic and schizotypal traits. To our knowledge, this study is the first to investigate differences in Bayesian inference along both trait spectra in a single task. Furthermore, this study is the first visual study to computationally disentangle and quantitatively assess the variations in individuals' likelihoods and priors. Surprisingly, schizotypal traits were found to have no effect on task performance and thus were not associated with any differences in the underlying statistical learning and Bayesian inference. For autistic traits, however, significant behavioral differences in prior integration were found, which were due to an increase in the precision of internal sensory representations in participants with higher AQ. Whether the current results extend to clinical populations will have to be examined in the future.

Methods

Participants

91 (47 females, 44 males, age range: 18-69) naïve participants with no motor disabilities and with normal (or corrected to normal) vision were recruited from the general population. We advertised for participants using posters and the internet across University of Edinburgh locations and other sites across Edinburgh. All participants gave informed written consent and received monetary compensation for participation. The study was approved by the University of Edinburgh School of Informatics Ethics Panel.

Questionnaires

ASD was assessed using 50-item version Autism Spectrum Quotient (AQ) [40], which is commonly used for assessing milder variants of autistic-like traits within the general population. Schizotypal traits were assessed using The Rust Inventory of Schizotypal Cognitions (RISC) [41]. RISC is specifically developed to measure schizotypal traits in the general population. In addition, a sub-group of 41 participants also completed Schizotypal Personality Questionnaire (SPQ) [42]. Finally, all participants were also asked to complete the Warwick-Edinburgh Mental Well-being Scale (WEMWBS)[43] in order to control for potential depression-induced differences in performance [44].

Apparatus

The visual stimuli were generated using Matlab Psychophysics Toolbox [45]. Participants viewed the display in a dark room at a distance of 80-100cm. The stimuli consisted of a cloud of dots with a density of 2 dots/deg^2 moving coherently (100%) at a speed of 9°/sec. Dots appeared within a circular annulus with minimum diameter of 2.2° and maximum diameter of 7°. The stimuli were displayed on a Dell P790 monitor running at 1024×768 at 100 Hz. The display luminance was calibrated using a Cambridge Research Systems Colorimeter (ColorCal MKII).

The task

The task was developed previously in our laboratory [29]. Participants have to: i) estimate the direction of coherently moving simple stimuli (dots) that are presented at low contrast levels (estimation task) and then ii) indicate whether they have actually perceived the stimulus or not (detection task). Since Chalk et al.[29] had shown that the effects of acquired priors become significant within the first 200 trials, instead of two experimental sessions of 850 trials each as in the original study, we used a single session of 567 trials (lasting around 40 min).

Each trial started by first displaying a fixation point (0.5°, 12.2 cd/m^2) for 400 ms, after which a field of moving dots appeared along with an orientation bar (length 1.1°, width 0.03°, luminance 4 cd/m^2, extending from the fixation point). Initial angle of the bar was randomized for each trial. Participants had to estimate the direction of motion by aligning the bar (using a computer mouse) to the direction the dots were moving in, and by clicking the mouse button to validate their estimate. The display cleared when either the participant had clicked the mouse or when 3000ms had elapsed. On trials where no stimulus was presented, the bar still appeared for the estimation task to be completed.

After a 200ms delay, the participants had to indicate whether they had actually detected the presence of dots in the estimation period (detection task). The display was divided into two parts by a vertical white line across the center of the screen, the left-hand side area reading "NO DOTS" and the right-hand side area reading "DOTS" (**Fig. 2a**). The cursor appeared in the center of the screen, and participants had to move it to the left or right and click to indicate their response. Immediate feedback for correct or incorrect detection responses was given by a cursor flashing

green or red, respectively. The screen was cleared for 400 ms before the start of a new trial. Every 20 trials, participants were presented with feedback on their estimation performance in terms of average estimation error in degrees (e.g., "In the last 20 trials, your average estimation error was 23°"). Every 170 trials (i.e. on three occasions) participants were given a chance to "have a short break to rest their eyes", in order to prevent fatigue. Participants clicked when they were ready to continue.

Design

The stimuli were presented at four different levels of contrast: 0 contrast (no-stimulus trials), 2 low levels contrasts and high contrast, randomly mixed across trials. There were 167 trials with no stimulus. The 2 low levels of contrast were determined using 4/1 and 2/1 staircases on detection performance [46]. There were 243 trials following the 4/1 staircase and 90 trials following the 2/1 staircase. The remaining 67 trials were at high contrast, which was set to 3.51 cd/m^2 above the background luminance.

For the two low contrast levels, there was a predetermined number of possible directions: 0°, ±16°, ±32°, ±48°, and ±64° with respect to a reference direction. The reference direction was randomized for each participant. For the 2/1 staircased contrasts, each predetermined motion direction was presented equally frequently. Unbeknownst to participants, stimuli at high and 4/1 staircase contrasts were presented more frequently at -32° and +32° motion directions, resulting in a bimodal probability distribution (**Fig. 1b**). For the 4/1 staircase contrast level, the dots were moving at ±32° in 173 (~70%) trials and in all the other predetermined motion directions in the remaining 70 (~30%) trials equally frequently. At the highest contrast level, 34 (~50%) trials had the dots moving at ±32° and the remaining 33 (~50%) trials were at random directions (i.e. not just the predetermined directions).

Data analysis

Responses on high contrast trials were used as a performance benchmark to ensure that participants were performing the task adequately. The predefined inclusion criteria were: 1) at least 80% detection and 2) less than 30° root mean squared error of estimations. 8 out of 91

participants failed to satisfy at least one of the criteria and were excluded from further analysis (**Appendix 1—Figure 1**).

Data analysis on the estimation of motion directions was performed on 4/1 and 2/1 staircased contrast levels only and only on trials where participants both validated their choice with a click within 3000 ms in the estimation part and clicked "DOTS" in the detection part. The first 170 trials of each session were excluded from the analysis, as this was the upper limit for the convergence of the staircases to stable contrast levels (**Appendix 1—Figure 2**).

After removing these trials, the luminance levels achieved by the 2/1 and 4/1 staircases were found to be considerably overlapping (**Appendix 1—Figure 2**). Therefore, the data for both of these contrast levels was combined for all further analysis.
To account for random estimations (either accidental or intentional) that participants made on some trials, we fitted each participant's estimation responses to the probability distribution:

$$(1-\alpha)\cdot V(\theta|\mu,\kappa) + \alpha, \qquad (2)$$

Where α is the proportion of trials in which participant makes random estimates, and $V(\theta|\mu,\kappa)$ is the probability density function for the estimated angle θ for von Mises (circular normal) distribution with the mean μ and precision κ. The parameters μ and κ of the von Mises distribution were determined by maximizing the likelihood of the distribution in Eq. (2) for each presented angle.

To analyze the distribution of estimations in no-stimulus trials, we constructed histograms of 16° size bins. These histograms were converted into probability distributions by normalizing over all motion directions. We analyzed the estimation distribution when participants reported seeing dots (clicked "DOTS") within no-stimulus trials. We interpreted these false alarms as a simple form of perceptual hallucination.

Modelling

Bayesian models

Bayesian models assume that participants combined a learned prior of the stimulus directions with their sensory evidence in a probabilistic manner. We first assume that participants make noisy sensory observations of the actual stimulus motion direction (θ_{act}), with a probability

$$p_{sens}(\theta_{sens}|\theta_{act}) = V(\theta_t, \kappa_{sens}). \quad (3)$$

where θ_t itself varies from trial to trial around θ_{act} according to $p(\theta_t|\theta_{act}) = V(\theta_{act}, \kappa_{sens})$. While participants cannot access the "true" prior, $p(\theta)$, directly, we hypothesized that they learned an approximation of this distribution, denoted $p_{exp}(\theta)$. This distribution was parameterized as the sum of two von Mises distributions, centered on motion directions θ_{exp} and $-\theta_{exp}$, and each with precision κ_{exp} :

$$p_{exp}(\theta) = 0.5\,[V(-\theta_{exp}, \kappa_{exp}) + V(\theta_{exp}, \kappa_{exp})] \quad (4)$$

Combining these via Bayes' rule gives a posterior probability that the stimulus is moving in a direction θ:

$$p_{post}(\theta|\theta_{sens}) \propto p_{exp}(\theta) \cdot p_{sens}(\theta_{sens}|\theta) \quad (5)$$

The perceived direction, θ_{perc}, was taken to be the mean of the posterior distribution (almost identical results would be obtained by using the maximum instead). Finally, we accounted for motor precision and a possibility of random estimates on some trials via:

$$p(\theta_{est}|\theta_{perc}) = (1-\alpha) \cdot V(\theta_{perc}, \kappa_m) + \alpha, \quad (6)$$

where α is the proportion of trials in which participants make random estimates and κ_m is the motor precision.

Increased exposure to some motion directions might not only give rise to prior expectations, but also

induce learning in the sensory likelihood function itself 47,52. Therefore, we fitted two more model variants: 'BAYES_var' where κ_{sens} varied with the stimulus direction (i.e. it took five different values for each of the angles: 0°, ±16°, ±32°, ±48°, and ±64°) and 'BAYES_varmin' where κ_{sens} was allowed to be different for ±32° but was the same for all other directions.

Response strategy models

We wanted to test whether task behavior might be better explained by simple behavioral strategies. This class of models assumed that on trials when participants were unsure about the presented motion direction, they made an estimation based solely on prior expectations, while on the remaining fraction of trials they made unbiased estimates based solely on sensory inputs. The first model, 'ADD1', assumed that estimations derived from prior expectations were simply sampled from a learnt expected distribution, $p_{exp}(\theta)$ (see Chalk et al.[29] and **Appendix 2**). The second model, 'ADD2', was just as 'ADD1' except when participants were unsure about the stimulus motion direction, instead of sampling from the complete learned probability distribution ranging from -180° to $+180^{\circ}$, they effectively truncated this distribution on a trial by trial basis and sampled from only one part of it, negative (-180° to 0°) or positive (0° to $+180^{\circ}$), depending on which side of the distribution the actual stimulus occurred (see Chalk et al, 2010 and SI). We also considered slight variations of the 'ADD1' and 'ADD2' models, denoted 'ADD1_m' and 'ADD2_m' respectively. These were identical to 'ADD1' and 'ADD2' except from setting $1/\kappa_{exp}$ to zero; that is, on trials when perceptual estimates were derived only from expectations, they were equal to the mode of the learnt distribution (i.e. no uncertainty).

Parameter estimation

We used performance in high contrast trials to estimate motor precision, κ_m, for each individual. We assumed that, for those trials, sensory uncertainty was close to zero. Motor precision was then determined by fitting estimation responses to the distribution in Eq. (2) by replacing μ with the actual motion direction, θ_{act}. The estimated motor precision was used in all subsequent model fitting as a fixed parameter. The rest of the free parameters were estimated by fitting the response data at the two low (staircased) contrast levels. For each model with a set of free parameters *M*, we computed the probability distribution $p(\theta_{est}|\theta_{act}; M)$ of making an estimate θ_{est} given the actual

stimulus direction θ_{act}. For the response strategy models, by definition, the $p(\theta_{est}|\theta_{act}; M)$ corresponds to average behavior in the task.

The parameters were estimated by maximizing the fit of the log likelihood function for the experimental data for each participant individually. The maximum likelihood was found using a simplex algorithm, using *fminsearchbnd* Matlab function. To avoid convergence at a local maximum we constructed a grid of initial κexp and κsens parameter values covering the range found in previous studies. We selected the resulting set of parameters that corresponded to the largest log-likelihood.

Model Comparison

To compare the model fits we used Bayesian Information Criterion (BIC), which approximates the log of model evidence [48] :

$$-2 \cdot log(P(D|M)) \approx BIC = -2 \cdot log(P(D|M, \hat{\Theta})) + k \cdot log(n), \quad (7)$$

where M is model, D is observed data and $P(D|M, \hat{\Theta})$ is the likelihood of generating the experimental data given the most likely set of parameters, $\hat{\Theta}$; k is the number of model parameters and n is the number of data points (or equivalently, the number of trials). BIC evaluates the model by how it fits the data by also penalizing for model complexity (number of parameters); lower BIC score indicates a better model.

Parameter recovery

To determine whether the BAYES model can distinguish the effects of strong likelihoods from those of weak priors [10, 11] and to evaluate the robustness of our methods, we performed parameter recovery. First, we generated 80 sets of parameters (i.e. 80 synthetic individuals) by randomly sampling each parameter from a Gaussian distribution centered on the mean value of each parameter found in our sample (40° for θ_{exp}, 15° for σ_{exp}, 10° for σ_{sens}, 0.06 for α and 10° for σ_{motor}). Second, for each set of parameters, we simulated data for 200 trials with the Bayesian model by randomly sampling from the estimation probability distribution. We used 200 simulated

trials only, to match the empirical data (200 corresponds to the amount of experimental trials used for fitting, after excluding high contrast and zero contrast trials).[1] Finally, we fitted the BAYES model to the simulated data. To evaluate the goodness of recovered parameters, we computed Pearson's correlation between the actual parameters and the recovered parameters.

Statistical tests

Due to the presence of outliers in many of the measures, we used robust regression techniques for measuring the presence and strength of the effects in our data. This was done using *robustfit* function in Matlab, which downweighs the influence of outliers in proportion to their distance from the regression line, which is computed via iteratively reweighted least squares (IRLS)[53]. For the loss function we used Huber function[54] with a tuning constant of 1.345, which corresponds to 95% estimator efficiency as compared to ordinary least squares.

Furthermore, we applied Bonferroni correction for multiple testing based on the number of independent hypotheses that we tested; that is, whether two personality traits, ASD and schizotypy, were associated with the two variables of interest, acquired priors and sensory likelihoods, - this resulted in 4 different hypotheses. Note that while the number of null hypothesis significance tests that we performed exceeds this number, the tests within each set concerning the same hypothesis were not independent (each test was based on derivative and/or correlated values to those in the other tests within the same set), and thus would not have met the independence assumption on which Bonferroni correction is based.

Finally, due to the limitations of frequentist statistics for accepting the null hypothesis, we performed Bayesian correlation analysis and computed Bayesian Factors[55] for the null hypothesis (BF_{01}). This was done using JASP[56] (Version 0.8.6). Due to the presence of outliers, this analysis was carried out using the non-parametric Kendall's Tau-b correlation coefficient.

[1] Simulating more trials would result in a better parameter recovery but the results would no longer be informative about the reliability of parameters estimated from empirical data.

Acknowledgements

We thank Gizem Aras for assisting in data collection, and Katie Richards for assisting with participants' recruitment.

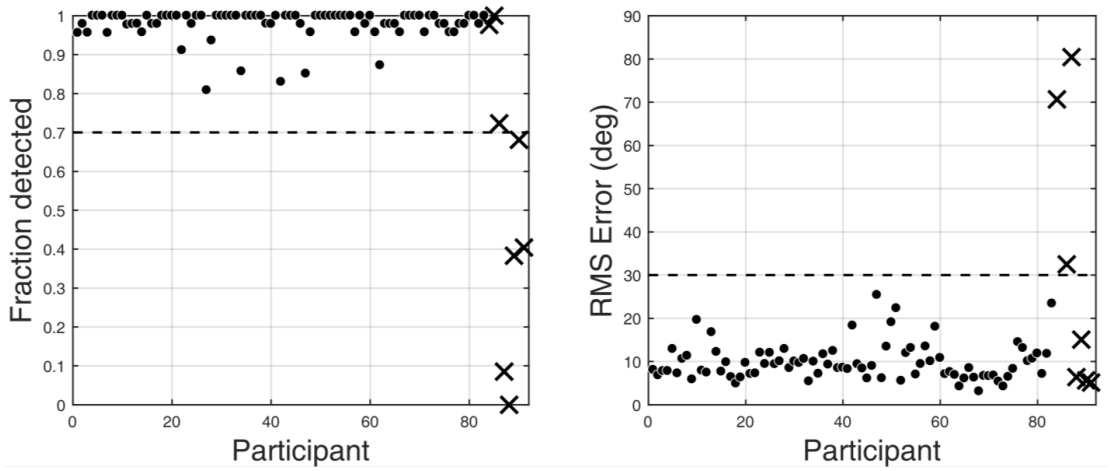
Competing financial interests

The authors declare no competing financial interests.

Appendix 1

Exclusion criteria

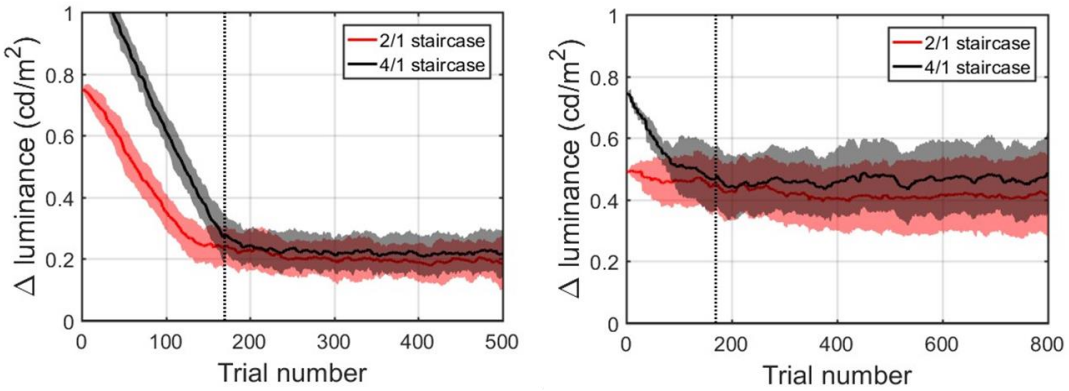
In order to ensure that participants performed adequately in the psychophysical task, we used predetermined performance criteria for inclusion into the study. Firstly, participants were required to detect the motion stimuli on more than 80% of trials with the high contrast motion stimuli and also make active estimates of the motion directions by clicking the mouse. Secondly, their average estimation performance on the high contrast stimuli had to be within 30° of the correct angle. 8 out of 91 participants failed to satisfy at least one of the criteria: 2 participants did not satisfy the first criteria, 4 did not satisfy the second criteria and 2 did not satisfy both of the criteria (**Appendix 1—Figure 1**). These participants were excluded from further analysis.



Appendix 1—Figure 1. Task performance at the highest contrast level and exclusion Criteria. Left panel: fraction of detected high contrast trials - quantified as the fraction of trials in which participants both validated their choice with a click within 3000 ms in the estimation part and reported seeing dots (clicked "DOTS") in the detection part. Right panel: root mean square error of estimations on high contrast trials. The dashed lines represent minimum performance criteria (more than 80% detection and less than 30° RMS error of estimations). Excluded participants are denoted by cross markers.

Staircased stimulus contrast levels

Appendix 1—Figure 2 describes the average convergence of the contrast staircases. Two groups comprising our sample performed the task at different background contrast levels. For a subgroup of 50 participants (left panel), the background luminance was set to 1.16 cd/m^2 for the other sub-group of 41 (right panel) it was set to 5.18 cd/m^2 For both groups, contrast staircases converged after 170 trials for both intermediate contrast levels, denoted with the vertical dashed line. In both groups, 2/1 and 4/1 staircased contrasts were considerably overlapping: on average 2/1 being 0.20±0.04 cd/m^2 and 4/1 being 0.22±0.04 cd/m^2 above the 1.16 cd/m^2 background luminance; and on average 2/1 being 0.42±0.05 cd/m^2 and 4/1 being 0.46±0.05 cd/m^2 above the 5.18 cd/m^2 background luminance. Thus, the two intermediate contrasts were combined for all further data analysis.



Appendix 1—Figure 2. Population averaged stimulus contrast relative to the background contrast for the 2/1 (red) and 4/1 (black) staircased contrast levels. Standard deviation is denoted by shaded areas with corresponding colors. The vertical dashed line marks 170 trials. Left panel: 44 participants (remaining after exclusion) that performed the task with the background luminance set to 1.16 cd/m^2. Right panel: 39 participants (remaining after exclusion) that performed the task with the background luminance set to 5.18 cd/m^2.

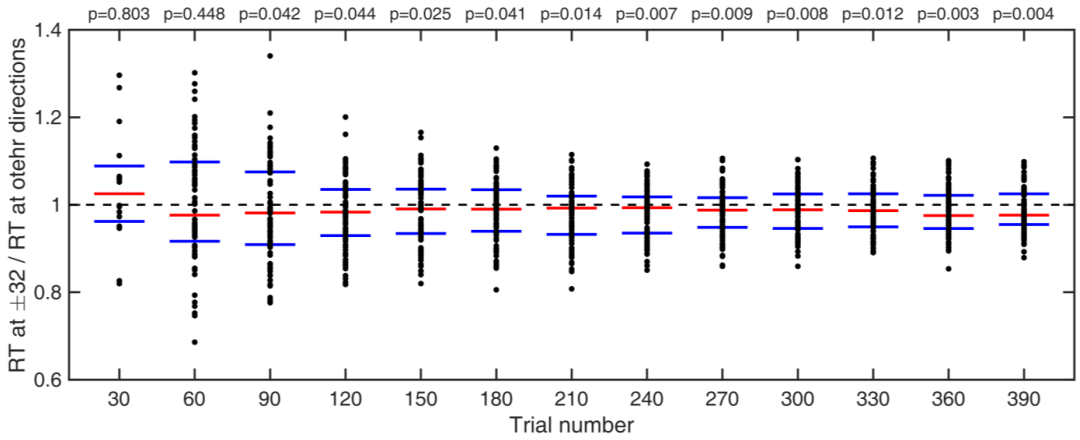
Combining the different background luminance levels

To compare the two sub-groups that performed the task at different background luminance levels, we performed Wilcoxon two-tailed rank sum test for all of the behavioral measures and none of them indicated any differences: mean absolute estimation bias (z = 0.652; ranksum = 1920; p =

0.514), mean variance of estimations (z = -0.406; ranksum = 1803; p = 0.685), total number of hallucinations (z = 0.128; ranksum = 1862; p = 0.898) number of hallucinations within 8° of ±32° (z = 0.870; ranksum = 1943; p = 0.384), mean estimation reaction time (z = 0.479; ranksum = 1901; p = 0.632). The two groups were therefore combined.

Temporal emergence of the impact of expectations

We investigated how many trials it took for the acquired prior effects to impact behavior. First, we looked at estimation reaction times (RT) and compared mean RT of each individual at ±32° with mean RT at all other directions; we compared cumulative moving averages at every 30 trials (**Appendix 1—Figure 3**). We found that it took less than 90 trials for RT at ±32° to become significantly shorter than average RT at all other directions (**Appendix 1—Figure 3** and p-values within).



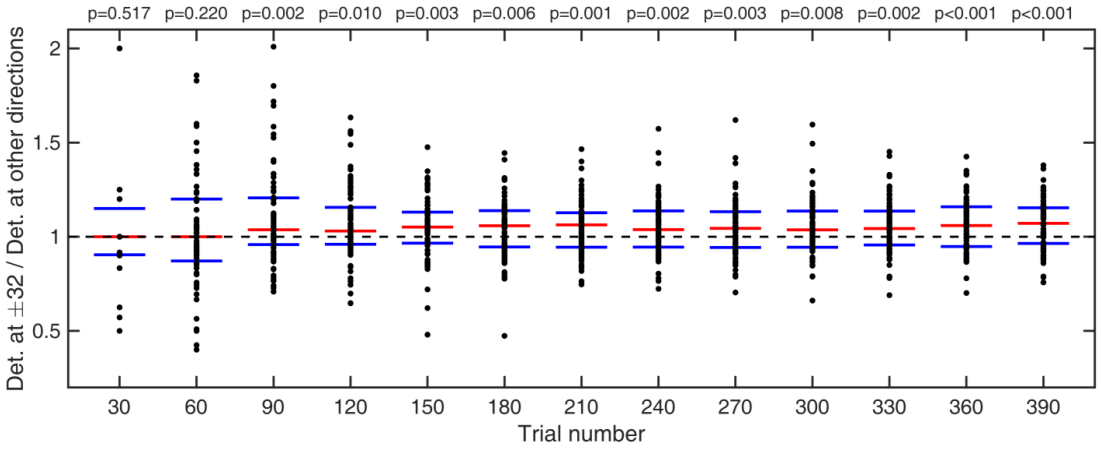
Appendix 1—Figure 3. Cumulative moving average of ratio of estimation reaction times at ±32° vs average reaction times at all other directions. Red bars indicate median values and blue bars indicate 25th and 75th percentiles. p-values indicate whether RTs at ±32° are significantly shorter than average RTs over all other directions (one-tailed Wilcoxon signed rank test).

Similarly, we looked at average detection performance and compared the fraction of trials in which stimulus was detected at ±32° with the mean fraction detected over all other presented directions; again, we compared cumulative moving averages at every 30 trials (**Appendix 1—Figure 4**). We found that it took less than 90 trials for detection at ±32° to become significantly better than average detection over all other presented directions (**Appendix 1—Figure 4** and p-values within).

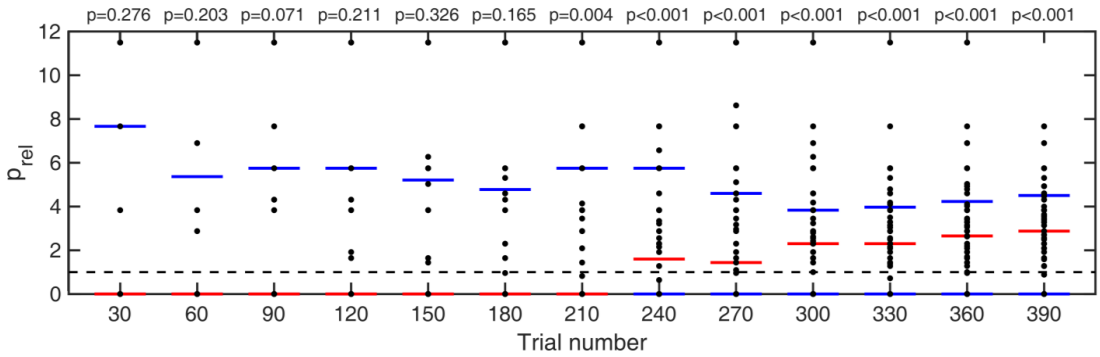
Lastly, for trials where no stimulus was presented, we looked at how long it took participants to start hallucinating predominantly around ±32° as opposed to all other possible directions. This was quantified as a probability ratio $p_{rel:}$

$$p_{rel} = p(\theta_{est} = \pm 32(\pm 8)°) \cdot N_{bins}\ , \qquad (1)$$

where N_{bins} is the number of bins (23), each of size 16°. This probability ratio would be equal to 1 if participants were equally likely to estimate within 8° of ±32° as they were to estimate within other bins. Again, we computed cumulative moving mean at every 30 trials (**Appendix 1—Figure 5**). For participants who did not report seeing dots at any direction within a given number of trials (i.e. zero total hallucinations) this probability ratio was undefined, therefore, those individuals were omitted from significance test at that point. We found that it took less than 210 trials for p_{rel} to become significantly larger than 1 (**Appendix 1—Figure 5** and p-values within).



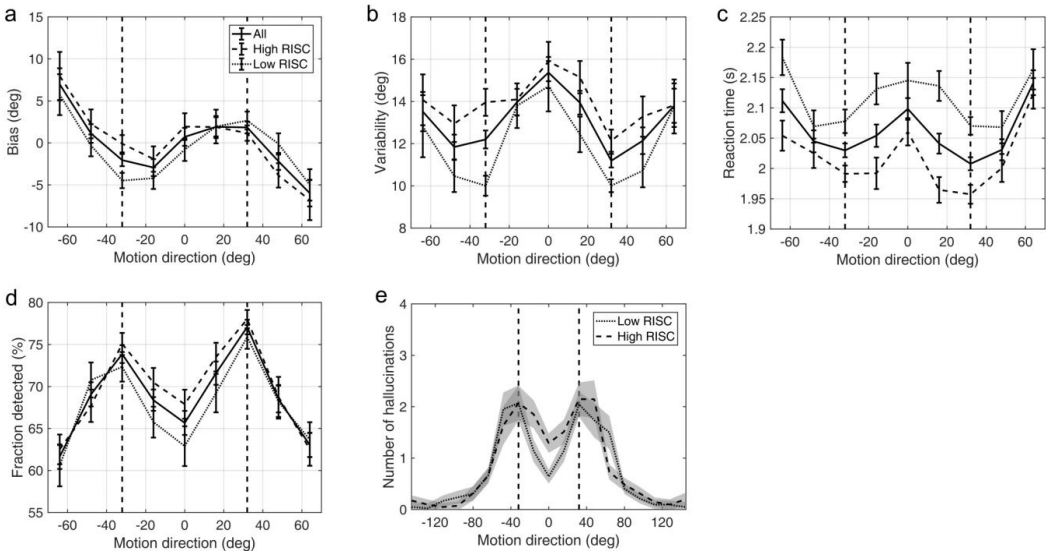
Appendix 1—Figure 4. Cumulative moving average of ratio of fraction of detected stimuli at ±32° vs average fraction detected at all other directions. Red bars indicate median values and blue bars indicate 25th and 75th percentiles. p-values indicate whether fraction detected at ±32° are significantly larger than average fraction detected over all other directions (one-tailed Wilcoxon signed rank test).



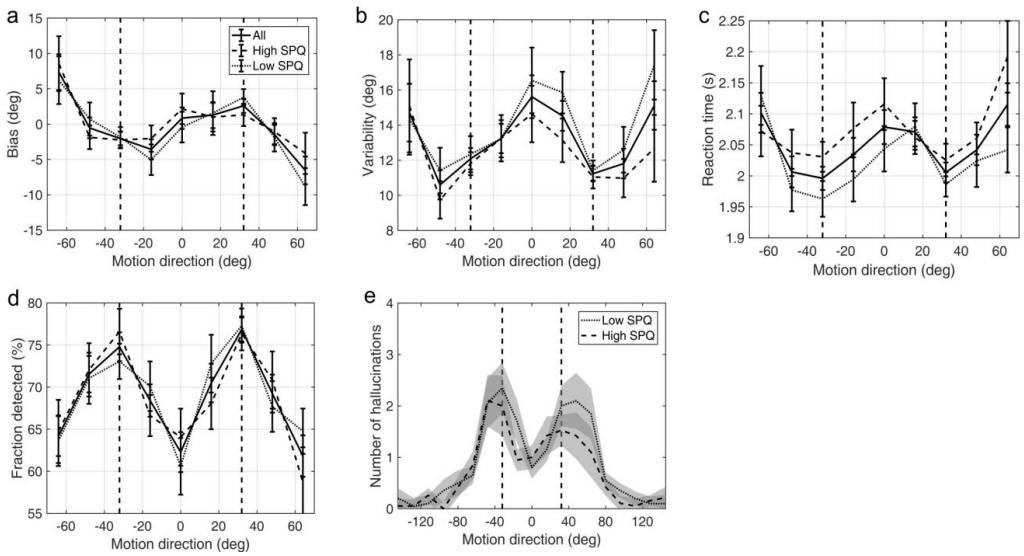
Appendix 1—Figure 5. Cumulative moving average of ratio of fraction of detected stimuli at ±32° vs average fraction detected at all other directions. Red bars indicate median values and blue bars indicate 25th and 75th percentiles. p-values indicate whether fraction detected at ±32° are significantly larger than average fraction detected over all other directions (one-tailed Wilcoxon signed rank test).

Schizotypy traits and task performance

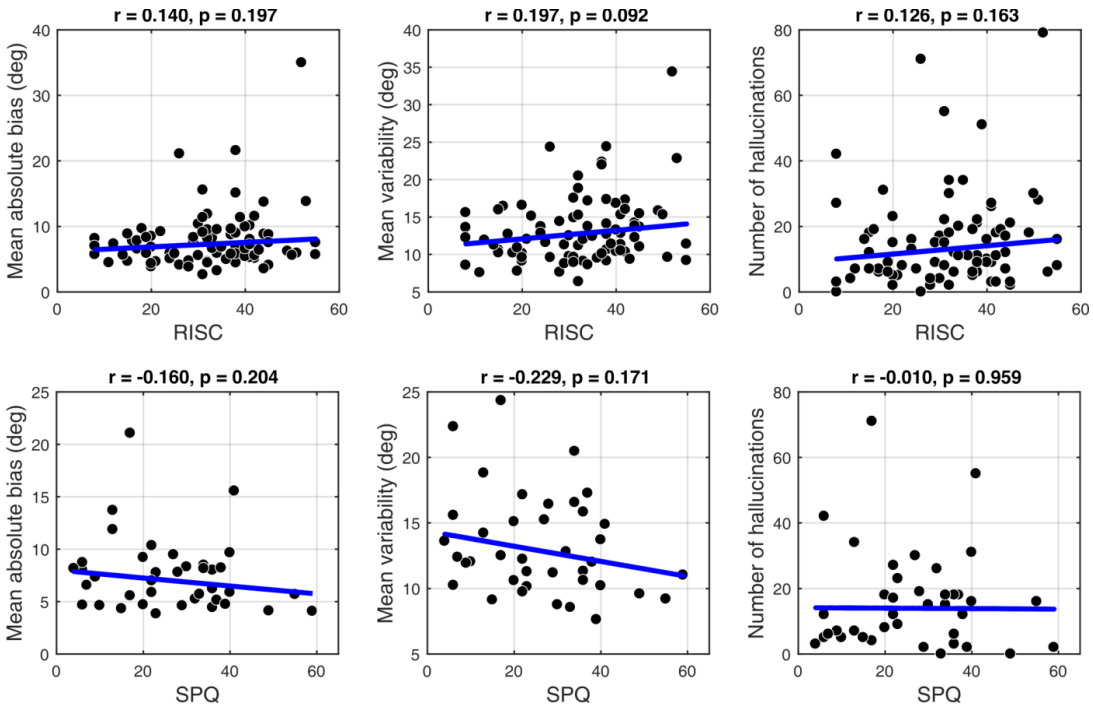
Appendix 1—Figure 6 and **Appendix 1—Figure 7** show task performance by groups which were formed by splitting the sample on the median RISC and SPQ scores respectively. **Appendix 1—Figure 8** shows the correlations between RISC and SPQ scores and the corresponding performance measures. There were no significant correlations with any of the measures.



Appendix 1—Figure 6. Average group performance on low-contrast trials (**a-d**) and on trials with no stimulus (**e**) by groups split by median RISC score. (**a**) Mean estimation bias, (**b**) standard deviation of estimations, (**c**) estimation reaction time and (**d**) fraction of trials in which the stimulus was detected. (**e**) Distribution of hallucinations. The vertical dashed lines correspond to the two most frequently presented motion directions (±32°). Error bars and shaded areas represent within-subject standard error.



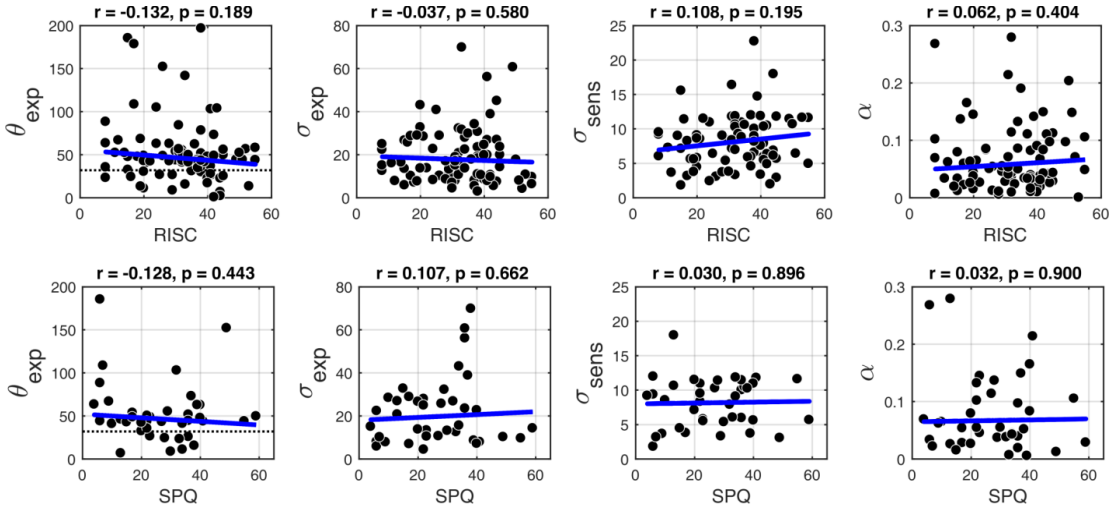
Appendix 1—Figure 7. Average group performance on low-contrast trials (**a-d**) and on trials with no stimulus (**e**) by groups split by median SPQ score. (**a**) Mean estimation bias, (**b**) standard deviation of estimations, (**c**) estimation reaction time and (**d**) fraction of trials in which the stimulus was detected. (**e**) Distribution of hallucinations. The vertical dashed lines correspond to the two most frequently presented motion directions (±32°). Error bars and shaded areas represent within-subject standard error.



Appendix 1—Figure 8. Correlations between personality traits, RISC (top row) and SPQ (bottom row) and task performance. There were no significant correlations with any of the measures: mean absolute bias (left column), mean estimation variability (middle column) and total number of hallucinations (right column). Robust correlation coefficients and p-values are indicated above each plot. The blue lines denote robust regression.

Schizotypy traits and model parameters

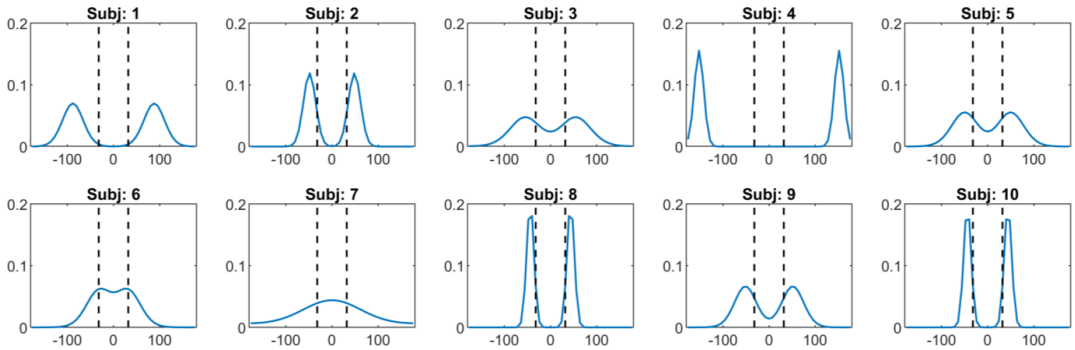
Appendix 1—Figure 9 shows the robust correlation analysis results between the BAYES model parameter estimates and schizotypy scores. There was no significant correlation with any of the parameters. Further Bayesian correlation analysis provided positive evidence that schizotypy traits had no effect on prior precision (RISC: τ_b = -0.012, BF_{01} = 6.90; SPQ: τ_b = 0.071, BF_{01} = 3.97).



Appendix 1—Figure 9. Correlations with the BAYES model parameter values and schizotypy traits (as measured by both RISC and SPQ). First column: θ_{exp} - mean of the prior expectations, second column: σ_{exp} - uncertainty of the prior distribution, third column: σ_{sens} - uncertainty in the sensory likelihood and fourth column: α - fraction of random estimations. Robust correlation coefficients and p-values are indicated above each plot. The blue lines denote robust regression.

Individual priors recovered via BAYES model

Appendix 1—Figure 10 shows a representative sample of the priors we extracted for a number of individuals, using the 'BAYES' model.



Appendix 1—Figure 10. A representative sample of prior expectations for each individual as reconstructed via 'BAYES' model. The dashed lines correspond to the two most frequently presented motion directions (±32°).

Appendix 2

Response bias models

We wanted to account for the possibility that the task behavior might be better explained by simple behavioral strategies. This class of models assumed that on trials when participants were unsure about the presented motion direction they made an estimation based solely on prior expectations, while on the remaining fraction of trials they made unbiased estimates based solely on sensory input.

ADD1

The first model ('ADD1') assumed that when participants were unsure about which motion direction they had perceived, they made an estimate that was close to one of the two most frequently presented motion directions. In this model, on each trial, participants make a sensory observation of the stimulus motion direction, θ_{obs}. We parameterize the probability of observing the stimulus to be moving in a direction θ_{obs} by a von Mises (circular normal) distribution centered on the actual stimulus direction and with width determined by $1/\kappa_{sens}$:

$$p_{sens}(\theta_{sens}|\theta_{act}) = V(\theta_{act}, \kappa_{sens}) \quad (3)$$

On most trials, we assume that participants make a perceptual estimate of the stimulus motion direction (θ_{perc}) that is based entirely on their sensory observation so that $\theta_{perc} = \theta_{obs}$. However, on a certain proportion of trials, when participants are uncertain about whether a stimulus was present or not, they resort to their expectations by making a perceptual estimate that is sampled from a learned distribution, $p_{exp}(\theta)$. For simplicity, we parameterize this distribution as the sum of two circular normal distributions, each with width determined by $1/\kappa_{exp}$, and centered on motion directions $-\theta_{exp}$ and θ_{exp}, respectively. Finally, we accommodate for the fact that there will be a certain amount of noise associated with moving the estimation bar to indicate which direction the stimulus is moving in as well as allowing for a fraction of trials α, where participants make estimates that are completely random. Thus, the estimation response θ_{est} is related to the perceptual estimate θ_{perc} via the equation:

$$p(\theta_{est}|\theta_{perc}) = (1-\alpha) * V(\theta_{perc}, \kappa_m) + \alpha. \quad (4)$$

Bringing all this together, the distribution of estimation responses for a single participant is given by:

$$p(\theta_{est}|\theta_{act}) = (1-\alpha)[(1-a(\theta))p_l(\theta_{obs} = \theta_{est}|\theta_{act}) + a(\theta)p_{exp}(\theta_{est})] * V(0, \kappa_m) + \alpha. \quad (5)$$

where the asterisk denotes a convolution and $a(\theta)$ determines the proportion of trials that participants sampled from the expected distribution, $p_{exp}(\theta)$. The resulting 'ADD1' model has 9 free parameters θ_{exp}, k_{exp}, $a(\theta)$ (which can take a different value for each of the 5 angles: 0°, ±16°, ±32°, ±48°, and ±64°), κ_{sens} and α.

ADD2

The second model, 'ADD2', was just as 'ADD1' except that it had slightly more complex strategy for trials when participants were unsure about the stimulus motion direction: instead of sampling from the complete learned probability distribution ranging from -180° to $+180^{\circ}$ (Eq. (11)), they effectively truncated this distribution on a trial by trial basis and sampled from only one part of it, negative (-180 to 0°) or positive (0 to $+180^{\circ}$), depending on which side of the distribution the actual stimulus occurred. Incorporating this into the distribution of estimation responses gives:

$$p(\theta_{est}|\theta_{act}) = (1-\alpha)[(1-a(\theta)-b(\theta))p_l(\theta_{obs} = \theta_{est}|\theta_{act}) + a(\theta)p_{expN}(\theta_{est}) + b(\theta)p_{expP}(\theta_{est})] * V(0, \kappa_m) + \alpha \quad (6)$$

where asterisk ($*$) denotes convolution; a(θ) and b(θ) determine the proportion of trials in which participants sample from either anticlockwise or clockwise distributions $p_{expN}(\theta)$ and $p_{expP}(\theta)$, respectively.

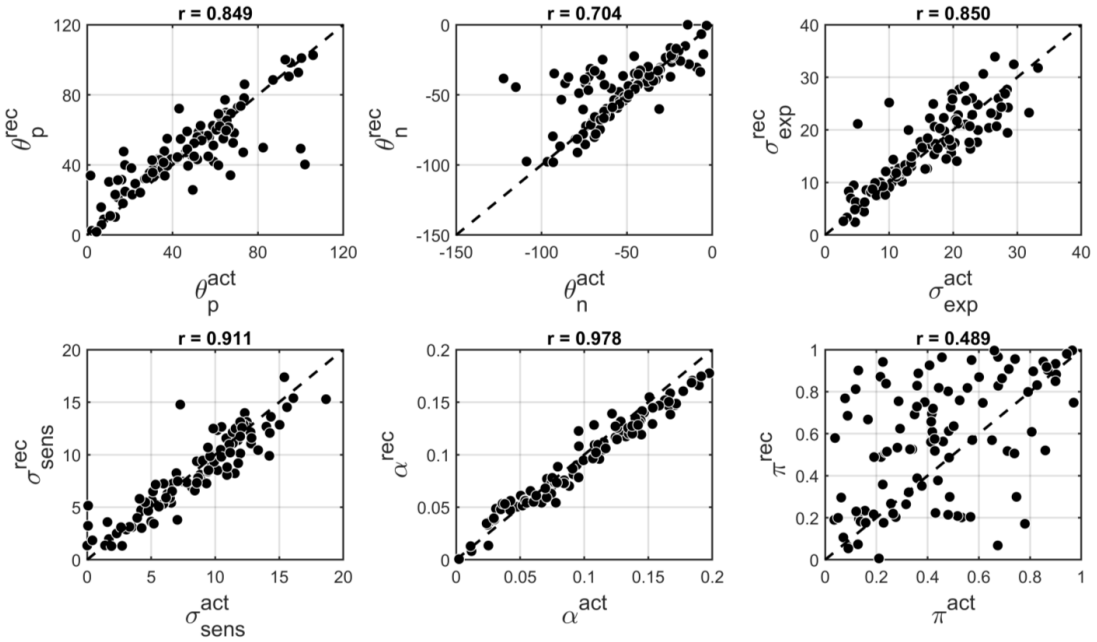
In addition, we also considered slight variations of the 'ADD1' and 'ADD2' models, denoted 'ADD1_m' and 'ADD2_m' respectively. These were identical to 'ADD1' and 'ADD2' except from setting $1/\kappa_{exp}$ to zero; that is, on trials when perceptual estimates were derived only from expectations, they were equal to the mode of the learnt distribution (i.e. no uncertainty).

Non-symmetric prior models

The stimulus distribution is multimodal and symmetric. Learning such a distribution might be inherently difficult. We reasoned that some individual differences might lie in asymmetries of the acquired priors. Therefore, we explored an alternative parameterization of the acquired priors which allowed them to be asymmetrical. We allowed the two modes in the prior to have different position with respect to 0° and to have different amount of probability associated with each mode. This resulted in:

$$p_{exp}(\theta) = (1 - \pi) \cdot V(\theta_p, \kappa_{exp}) + \pi \cdot V(\theta_n, \kappa_{exp}) \quad (2)$$

where π ($\in$ [0 1]) is a mixing parameter. Using this parameterization we fitted 'BAYES' model as described in the main text (thus, we denoted this alternative model as 'BAYES_π'). The alternative parameterization did not result in a better BIC as compared to 'BAYES' model ($p = 0.378$, signed rank test). In addition, we performed parameter recovery to determine how robust 'BAYES_π' is and found that recovering the mixing parameter π was not very reliable ($r=0.4$), although other parameters retained most of their previous reliability (**Appendix 2—Figure 1**). We thus focused on the simpler model in the current study.



Appendix 2—Figure 1. Comparison of actual and recovered parameters via 'BAYES_π' model. θ_p and θ_n - positive and negative modes of the bimodal distribution of prior expectations, σ_{exp} - uncertainty of the prior distribution, σ_{sens} uncertainty in the sensory likelihood, α - fraction of random estimations, π - mixing parameter responsible for the degree of bimodality. Actual parameters are scattered along x-axis and recovered parameters are scattered along y-axis. The dashed diagonal line is a reference line indicating perfect parameter recovery. Pearson's correlation coefficients are indicated above each plot.

Full models (estimation + detection)

We have built a Bayesian model that incorporates both estimation and detection performance ('BAYES_full') in order to fully account for the task behavior. This time, the acquired priors consisted of both the expectations about the direction of stimuli motion (θ) and the expectations about whether stimulus is presented (s=1) or not (s=0). It was parameterized as:

$$p_{exp}(\theta, s) = \begin{cases} (1-b)\cdot\frac{1}{2\pi}, & \text{if } s = 0 \\ b\cdot\frac{1}{2}[V(-\theta_{exp}, \kappa_{exp}) + V(\theta_{exp}, \kappa_{exp})], & \text{if } s = 1 \end{cases}$$

where parameter *b* accounts for a participant's average expectation that the stimulus will be presented. Thus, we assumed that expectations about motion direction were uniform for when no stimulus was expected. While the expectations about motion direction when the stimulus was expected followed the bimodal probability distribution just as in the previous models.

On each trial, given the presented motion direction (θ_{act}) and the presence of the stimulus (s), participants made sensory measurements $p_{sens}(\theta_{sens},s_{sens}|\theta_{act},s)$. For simplicity, we assumed that the sensory probability of whether the stimulus was present ($p_{sens}(s_{sens}|\theta_{act},s)$) was independent of the sensory input about the motion direction ($p_{sens}(\theta_{sens}|\theta_{act},s)$). We further assumed that s_{sens} was independent of the presented motion direction θ_{act}, as informed by 'BAYES_var' model (that allowed the sensory likelihood to vary based on the presented motion direction), which did not produce a better fit. As before, the mean of the motion direction was allowed to fluctuate on trial-by-trial basis, such that:

$$p(\theta|\theta_{act}) = V(\theta_{act},\kappa_{sens}) , \tag{7}$$

where κ_{sens} is sensory precision. Given the estimate of the mean θ, the sensory input θ_{sens} is represented with the associated uncertainty via:

$$p_{sens}(\theta_{sens}|\theta) = V(\theta,\kappa_{sens}) . \tag{8}$$

Putting all this together, the sensory likelihood was expressed as:

$$p_{sens}(\theta_{sens},s_{sens}|\theta,s) = p_{sens}(\theta_{sens}|\theta,s)p(s_{sens}|s) , \tag{9}$$

where $p_{sens}(\theta_{sens}|\theta_{act},s)$ was parameterized as:

$$p_{sens}(\theta_{sens}|\theta_{act},s) = \begin{cases} \frac{1}{2\pi}, & \text{if } s = 0 \\ V(\theta,\kappa_{sens}), & \text{if } s = 1 \end{cases}$$

where we assumed that sensory likelihood is uniform when no stimulus is presented. Finally, $p_{sens}(s_{sens}|s)$ was parameterized as:

$$p_{sens}(s_{sens} = \{0,1\}|s) = \begin{cases} \{1-c,c\}, & \text{if } s = 0 \\ \{1-d,d\}, & \text{if } s = 1 \end{cases}$$

where parameter *c* is the average probability of detecting dots when they are not presented, and parameter 'd' is the average probability of detecting dots when they are presented. Putting together prior and likelihood, the resulting posterior probability distribution becomes:

$$p_{post}(\theta,s|\theta_{sens},s_{sens}) \propto p_{sens}(\theta_{sens}|\theta,s) \cdot p_{sens}(s_{sens}|s) \cdot p_{exp}(\theta,s) . \tag{10}$$

With a given posterior participants could have performed detection task at least in two ways. One way is to maximize the posterior (i.e. to always choose the value of s that has higher probability):

$$s_{perc} = \text{argmax}\,[p_{post}(s|\theta_{sens}, s_{sens})] \quad (11)$$

Another way is to perform probability matching and choose in accordance to the size of the probabilities:

$$s_{perc} = \begin{cases} 0\ , & \text{if } p_{post}(s=0|\theta_{sens}, s_{sens}) > \eta \\ 1\ , & \text{if } p_{post}(s=0|\theta_{sens}, s_{sens}) < \eta \end{cases}$$

where $\eta \in [0\ 1]$ and is drawn for each trial from a uniform distribution. We considered both of these possibilities and implemented a variant of the model for each. Finally, just as in 'BAYES' model, the motion direction percept was formed by taking the mean of the posterior:

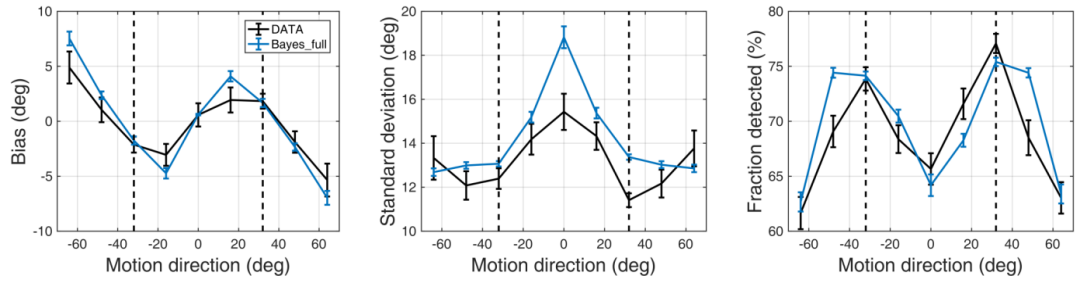
$$\theta_{perc} = \int \theta \cdot p_{post}(\theta|\theta_{sens}, s_{sens})\ d\theta\ = \frac{1}{Z}\int \theta \cdot \sum_s p_{exp}(\theta) \cdot p_{sens}(\theta_{sens}|\theta, s) \cdot p_{sens}(s_{sens}|s)\ d\theta\ , \quad (12)$$

As previously, we accounted for motor precision and the lapse responses via:

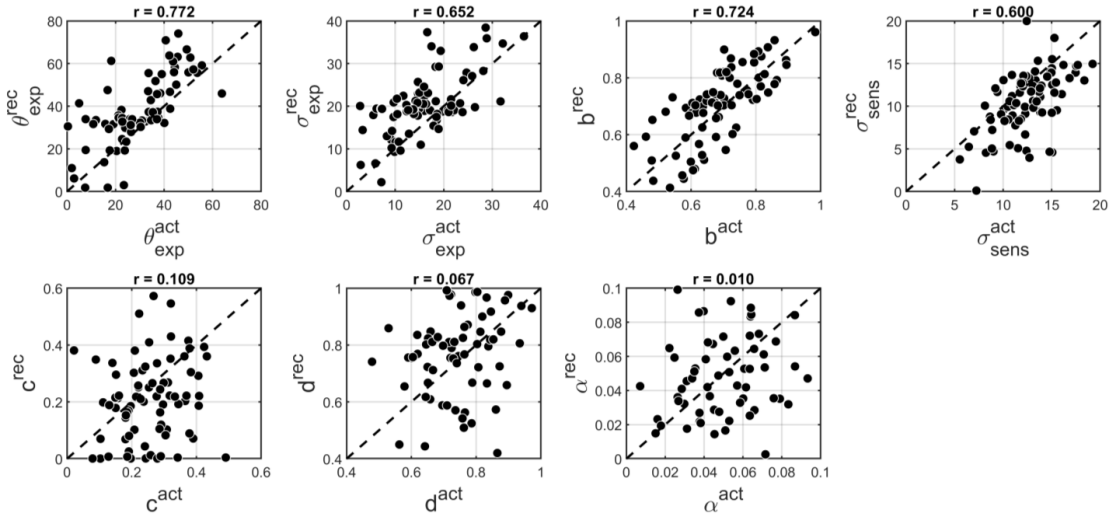
$$p(\theta_{est}|\theta_{perc}) = (1-\alpha) \cdot V(\theta_{perc}, \kappa_{motor}) + \alpha \cdot p_{exp}(\theta) * V(0, \kappa_{motor})\ . \quad (13)$$

In total, 'BAYES_full' model had 7 free parameters. To fit the model, in addition to intermediate contrast trials, we also used no-stimulus trial data. The rest of the fitting procedure was the same as in the main text: we built a distribution of 1,000 posterior estimations for each presented angle and one more distribution of 1,000 posterior estimations for no stimulus trials.

We found that 'BAYES_full' provided a good fit and captured the main features of both estimation and detection performance (Appendix 2—Figure 2). As before, to test how reliable parameters estimated for 'BAYES_full' model are, we performed parameter recovery. Just as for 'BAYES' parameter recovery described in the main text, we generated 80 sets of parameters and simulated 200 trials of data with 'BAYES_full' model for each of them. Then we fitted 'BAYES_full' to the simulated data. The results revealed that parameters 'd' and 'c' had very poor recovery (**Appendix 2—Figure 3**). We thus focused on the simpler model in the current study.



Appendix 2—Figure 2. Task performance as predicted by the BAYES_full model. Left panel: mean estimation bias at different motion directions. Middle panel: standard deviation of estimations at different motion directions. Right panel: fraction of detected stimuli at different motion directions. The dashed lines correspond to the two most frequently presented motion directions (±32°). Error bars represent within-subject standard error.



Appendix 2—Figure 3. Comparison of actual and recovered parameters via 'BAYES_full' model. θ_{exp} - the mean of prior expectations of motion direction, σ_{exp} - uncertainty of the prior expectations of motion direction, σ_{sens} - uncertainty in the sensory likelihood, α - fraction of random estimations, b - prior expectation for dots being presented, c likelihood of detecting the dots when they are not presented, d - likelihood of detecting the dots when they are presented. Actual parameters are scattered along x-axis and recovered parameters are scattered along y-axis. The dashed diagonal line is a reference line indicating perfect parameter recovery.

References

1. Fletcher, P. C. and Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10(1): 48–58.
2. Corlett, P., Frith, C. D., and Fletcher, P. (2009). From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology*, 206(4):515–530.
3. Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., and Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in psychiatry*, 4.
4. Pellicano, E. and Burr, D. (2012). When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends in cognitive sciences*, 16(10): 504–510.
5. Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de Wit, L., and Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychological Review*, 121(4):649.
6. Lawson, R. P., Rees, G., and Friston, K. J. (2014). An aberrant precision account of autism. *Front Hum Neurosci*, 8.
7. Palmer, C. J., Lawson, R. P., and Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin* 143(5): 521-542.
8. Notredame, C.-E., Pins, D., Deneve, S., and Jardri, R. (2014). What visual illusions teach us about schizophrenia. *Frontiers in Integrative Neuroscience*, 8.
9. Speechley, W. J., Whitman, J. C., and Woodward, T. S. (2010). The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. *Journal of psychiatry & neuroscience: JPN*, 35(1): 7.
10. Brock, J. (2012). Alternative Bayesian accounts of autistic perception: comment on Pellicano and Burr. *Trends in cognitive sciences*, 16(12): 573–574.
11. Pellicano, E. and Burr, D. (2012). Response to Brock: noise and autism. *Trends in cognitive sciences*, 16(12):574–575.
12. Van Boxtel, J. J. and Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in psychology*, 4.
13. Van de Cruys, S., de Wit, L., Evers, K., Boets, B., and Wagemans, J. (2013). Weak priors versus overfitting of predictions in autism: Reply to Pellicano and Burr (tics, 2012). *I-Perception*, 4(2):95–97.
14. Powell, G., Meredith, Z., McMillin, R., and Freeman, T. C. (2016). Bayesian models of individual differences: Combining autistic traits and sensory thresholds to predict motion perception. *Psychological science*, 27(12): 1562–1572.
15. Skewes, J. C., Jegindø, E.-M., and Gebauer, L. (2014). Perceptual inference and autistic traits. *Autism*, 19(3):301–307.
16. Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., Goodyer, I. M., and Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis

and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences*, 112(43):13401–13406.

17. Schmack, K., de Castro, A. G.-C., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J.-D., Heinz, A., Petrovic, P., and Sterzer, P. (2013). Delusions and the role of beliefs in perceptual inference. *Journal of Neuroscience*, 33(34): 13701– 13712.
18. Nelson, M., Seal, M., Pantelis, C., and Phillips, L. (2013). Evidence of a dimensional relationship between schizotypy and schizophrenia: a systematic review. *Neuroscience & Biobehavioral Reviews*, 37(3): 317–327.
19. Van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P., and Krabbendam, L. (2009). A systematic review and meta- analysis of the psychosis continuum: evidence for a psychosis proneness–persistence–impairment model of psychotic disorder. *Psychological medicine*, 39(2): 179–195.
20. Constantino, J. N. and Todd, R. D. (2003). Autistic traits in the general population: a twin study. *Archives of general psychiatry*, 60(5): 524–530.
21. Karaminis, T., Cicchini, G. M., Neil, L., Cappagli, G., Aagten-Murphy, D., Burr, D., and Pellicano, E. (2016). Central tendency effects in time interval reproduction in autism. *Scientific reports*, 6.
22. Skewes, J. C. and Gebauer, L. (2016). Brief report: Suboptimal auditory localization in autism spectrum disorder: Support for the Bayesian account of sensory symptoms. *Journal of autism and developmental disorders*, 46(7): 2539– 2547.
23. Pell, P. J., Mareschal, I., Calder, A. J., von dem Hagen, E. A., Clifford, C. W., Baron-Cohen, S., and Ewbank, M. P. (2016). Intact priors for gaze direction in adults with high-functioning autism spectrum conditions. *Molecular autism*, 7(1): 25.
24. Croydon, A., Karaminis, T., Neil, L., Burr, D., and Pellicano, E. (2017). The light-from-above prior is intact in autistic children. *Journal of Experimental Child Psychology*, 161: 113–125.
25. Manning, C., Kilner, J., Neil, L., Karaminis, T., and Pellicano, E. (2016). Children on the autism spectrum update their behaviour in response to a volatile environment. *Developmental science*.
26. Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*, 20(9), 1293-1299.
27. Schmack, K., Schnack, A., Priller, J., and Sterzer, P. (2015). Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. *Schizophrenia Research: Cognition*, 2(2): 72–77.
28. Schmack, K., Rothkirch, M., Priller, J., and Sterzer, P. (2017). Enhanced predictive signalling in schizophrenia. *Human brain mapping*, 38(4): 1767–1779.
29. Chalk, M., Seitz, A. R., and Seriès, P. (2010). Rapidly learned stimulus expectations alter perception of motion. *Journal of Vision*, 10(8): 2–2.
30. Palminteri, S., Wyart, V., and Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences*, 21(6): 425-433.
31. Powers, A. R., Mathys, C. and Corlett, P.R. (2017). Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science* 357(6351): 596-600.

32. Seriès, P. and Seitz, A. R. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience*, 7:668.
33. Dickinson, A., Jones, M., and Milne, E. (2014). Oblique orientation discrimination thresholds are superior in those with a high level of autistic traits. *Journal of Autism and Developmental Disorders*, 44(11): 2844.
34. Bertone, A., Mottron, L., Jelenic, P., and Faubert, J. (2005). Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. *Brain*, 128(10): 2430–2441.
35. Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11): 1432–1438.
36. Bertone, A., Mottron, L., Jelenic, P., and Faubert, J. (2003). Motion perception in autism: a "complex" issue. *Journal of cognitive neuroscience*, 15(2):218–225.
37. Happé, F. and Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, 36(1): 5–25.
38. Plaisted, K. C. (2015). Reduced generalization in autism: an alternative to weak central coherence.
39. Harris, H., Israeli, D., Minshew, N., Bonneh, Y., Heeger, D. J., Behrmann, M., and Sagi, D. (2015). Perceptual learning in autism: over-specificity and possible remedies. *Nature Neuroscience*, 18(11): 1574.
40. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of autism and developmental disorders*, 31(1): 5–17.
41. Rust, J. (1988). The Rust inventory of schizotypal cognitions (RISC). *Schizophrenia Bulletin*, 14(2): 317.
42. Raine, A. (1991). The SPQ: a scale for the assessment of schizotypal personality based on DSM-III-r criteria. *Schizophrenia bulletin*, 17(4): 555.
43. Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., and Stewart-Brown, S. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. *Health and Quality of life Outcomes*, 5(1): 1.
44. Austin, M.-P., Mitchell, P., and Goodwin, G. M. (2001). Cognitive deficits in depression. *The British Journal of Psychiatry*, 178(3): 200–206.
45. Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10:433–436.
46. Garcıa-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision research*, 38(12): 1861–1881.
47. Stocker, A. A. and Simoncelli, E. P. (2006). Sensory adaptation within a Bayesian framework for perception. In *Advances in neural information processing systems*, 1289–1296.
48. Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, 33(2): 261–304.
49. Hemsley, D. R., & Garety, P. A. (1986). The formation of maintenance of delusions: a Bayesian analysis. The British Journal of Psychiatry, 149(1), 51-56.

50. Friston, K. (2005). A theory of cortical responses. Philosophical transactions of the Royal Society B: Biological sciences, 360(1456), 815-836.
51. Stephan, K. E., Baldeweg, T., & Friston, K. J. (2006). Synaptic plasticity and dysconnection in schizophrenia. Biological psychiatry, 59(10), 929-939.
52. Sato, Y., & Kording, K. P. (2014). How much to trust the senses: Likelihood learning. Journal of Vision, 14(13), 13-13.
53. Holland, P. W., & Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. Communications in Statistics-theory and Methods, 6(9), 813-827.
54. Huber, P. J. (1964). Robust estimation of a location parameter. The annals of mathematical statistics, 73-101.
55. Kass, R. E., & Raftery, A. E. (1995). Bayes factors. Journal of the american statistical association, 90(430), 773-795.
56. JASP Team. (2017). JASP (Version 0.8.6). https://jasp-stats.org/faq/how-do-i-cite-jasp/.
57. Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. Molecular autism, 6(1), 2.

3

Perceptual Bayesian inference in people with SCZ diagnosis

This chapter includes a postprint of a published journal article: Valton, V., Karvelis, P., Richards, K. L., Seitz, A. R., Lawrie, S. M., Seriès, P. (2019). Acquisition of visual priors and induced hallucinations in chronic schizophrenia. Brain. 142(8), 2523-2537.

Vincent Valton and I co-first authored this paper as this project had been started during Vincent's PhD - a preliminary analysis can be found in his thesis (Valton, 2014). Since then, the project has been advanced in many ways. Firstly, the sample size has gone up from 10 patients and 10 controls to 20 patients and 23 controls. Secondly, the current analysis employs more sophisticated statistical methods: mixed ANOVA for simultaneously testing behavioral task effects and group differences, Bayes factors for testing the null hypothesis and random effect Bayesian model selection for model comparison. Thirdly, the current analysis features a new winning computational model (a Bayesian model where lapse estimations follow the acquired prior instead of being random), which provided a unified explanation for group differences in different task conditions. Finally, the

current analysis also includes parameter recovery of the winning model, which demonstrates the reliability of the modelling results. All of the current analysis and results were produced by me (except for power analysis and Supplementary Figure 6, which were done by Vincent); however, they were accompanied by multiple discussions and decisions made jointly by me, Vincent and the other authors. The initial version of the paper was written by Vincent when reporting the preliminary analysis of the results in 2014, while I updated the manuscript to include all the new methods, results and recent relevant studies.

Acquisition of visual priors and induced hallucinations in chronic schizophrenia

Vincent Valton[1,2,5,¶], Povilas Karvelis[1,¶], Katie L. Richards[2], Aaron R. Seitz[3], Stephen M. Lawrie[2,4], Peggy Seriès[1,*]

[1] *Institute for Adaptive and Neural Computation, University of Edinburgh, UK*

[2] *Department of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, UK*

[3] *Department of Psychology, University of California Riverside, CA, USA*

[4] *Patrick Wild Centre, University of Edinburgh, Edinburgh, UK*

[5] *Institute of Cognitive Neuroscience, University College London, London, UK*

**Correspondence should be addressed to: pseries@inf.ed.ac.uk*

¶ These authors contributed equally to this work

Running Title: "Learning and Bayesian inference in schizophrenia"

Abstract

Prominent theories suggest that symptoms of schizophrenia stem from learning deficiencies resulting in distorted internal models of the world. To further test these theories, we here use a visual statistical learning task known to induce rapid implicit learning of the stimulus statistics (Chalk *et al.*, 2010). In this task, participants are presented with a field of coherently moving dots and need to report the presented direction of the dots (estimation task) and whether they saw any dots or not (detection task). Two of the directions were more frequently presented than the others. In controls, the implicit acquisition of the stimuli statistics influences their perception in two ways: 1- motion directions are perceived as being more similar to the most frequently presented directions than they really are (estimation biases); 2- in the absence of stimuli, participants sometimes report perceiving the most frequently presented directions (a form of hallucinations). Such behaviour is consistent with probabilistic inference, i.e. combining learnt perceptual priors with sensory evidence. We investigated whether patients with chronic, stable, treated schizophrenia (n=20) differ from controls (n=23) in the acquisition of the perceptual priors and/or their influence on perception. We found that, although patients were slower than controls, they showed comparable acquisition of perceptual priors, correctly approximating the stimulus statistics. This suggests that patients have no statistical learning deficits in our task. This may reflect our patients relative wellbeing on antipsychotic medication. Intriguingly, however, patients made significantly fewer hallucinations of the most frequently presented directions than controls and fewer prior-based lapse estimations. This suggests that prior expectations had less influence on patients' perception than on controls when stimuli were absent or below perceptual threshold.

Keywords: Schizophrenia, Inference, Statistical Learning, Hallucinations

Introduction

An increasingly popular idea in neuroscience is that perception and decision-making can be well described in terms of probabilistic inference processes (Knill and Pouget, 2004; Fiser *et al.*, 2010; Friston, 2010; 2012). For example, statistical and perceptual learning studies show that the perceptual systems continuously extract and learn statistical regularities of the environment (for a review in visual perception, see e.g. Seriès and Seitz, 2013). This learning results in the construction of internal models of the environment, or expectations, which are used automatically and unconsciously to predict and disambiguate perceptual inputs in situations of uncertainty and to guide decisions.

In this context, it has been proposed that psychiatric disorders in general, and schizophrenia in particular, might be explained in terms of deficits in probabilistic inference (Friston, 2005; Frith and Friston, 2012; Adams *et al.*, 2013; Jardri and Denève, 2013, for reviews see: Friston *et al.* 2016; Valton *et al.*, 2017; Sterzer et al 2018). Impaired statistical learning and/or inference deficits would lead to distorted internal models of the world, which could then explain the existence of abnormal beliefs or delusions experienced by patients with schizophrenia. Incorrect perceptual inference may also lead to severe forms of illusions and result in the complex hallucinations that are part of the positive symptoms of schizophrenia.

Two different lines of research support this general idea. First, a number of studies using explicit probabilistic learning tasks, such as the "beads task" report that patients with schizophrenia show a deficit in integrating probabilistic information resulting in faster responses than control subjects, an effect called the 'jumping-to-conclusions' bias (Huq *et al*., 1988; Speechley *et al*., 2010; Averbeck *et al.*, 2011; Evans *et al.*, 2012). Interestingly, patients with stronger delusional symptoms fare worse at the task than those who do not (Huq *et al*., 1988; Speechley *et al*., 2010), and control subjects displaying delusional ideation also show similar impairments at the task (Freeman *et al.*, 2008), suggesting a link between delusions and probabilistic inference (Garety *et al.*, 2013; Garety and Freeman 2013). Second, patients with schizophrenia do not experience visual illusions in the same way as controls do. For example, patients are less susceptible to certain visual illusions such as the hollow-mask illusion (Dima

et al., 2009; 2010; Keane *et al.*, 2013; for review see: Silverstein and Keane 2011a; 2011b; Notredame *et al.*, 2014). This suggests that they either have different implicit expectations about the environment (i.e. they would not have such a strong expectation that faces are convex) or that these expectations do not affect patients' perception in the same way as observed in controls.

A few studies have recently tried to test the impaired Bayesian inference hypothesis more directly, but the findings are mixed. Teufel et al. (2015) found that early psychosis and schizotypal traits were associated with an increased influence of prior knowledge when disambiguating two-tone images. Powers et al. (2017) also reported an increased perceptual prior influence in experimentally-induced hallucinations in both patients and controls with higher propensity to hallucinatory experiences. Interestingly, however, a series of studies by Schmack et al. (2013; 2015; 2017) found an increased influence of cognitive priors on the perception of bi-stable stimuli in participants with schizotypal traits and clinical schizophrenia, but showed on the contrary a decreased influence of perceptual priors, suggesting that the level at which the prior operates in the inference hierarchy might lead to differential effects. Finally, Jardri et al. (2017) investigated probabilistic reasoning and found schizophrenia patients to be over-counting sensory evidence and under-weighting priors, which they described using a circular inference model. Together these findings paint a complicated picture of Bayesian inference in schizophrenia where priors can have either increased or decreased influence depending on the task, the stimulus and the type of priors involved (e.g. low-level perceptual prior *vs*. high-level cognitive prior).

A general limitation of these studies however, is that it is typically unclear to what extent the effects are driven by deficits in statistical learning (i.e. forming and updating the priors) or impaired inference *per se*. Moreover, past studies were usually only qualitatively comparing the behavioural results they collected with the proposed Bayesian theories.

To address these issues, we simultaneously investigated the implicit acquisition of priors, how these priors are integrated with sensory information and the influence they have on perception when stimulus is absent (i.e. experimentally induced hallucinations) in patients with schizophrenia. We used a previously developed statistical learning task (Chalk *et al.*, 2010;

Gekas *et al.*, 2013; Karvelis *et al.*, 2018) that is known to induce the rapid acquisition of the statistics of motion stimuli. In this task, participants need to report the direction of motion of a cloud of dots (estimation task) and whether they have perceived the dots or not (detection task; on some trials no stimulus is presented). Unbeknownst to the participants, two directions of motion are more frequently presented than others. Participants implicitly and unconsciously learn those stimulus statistics. This learning influences perception such that: 1) motion stimuli are perceived as being more similar to the most frequently presented stimuli than they really are (i.e. estimation biases); 2) participants sometimes report perceiving the most frequently presented stimuli in absence of visual stimuli (a form of hallucination). In previous work (Chalk *et al* (2010) Karvelis et al (2018)), we showed that Bayesian modelling could be applied to individual participants' performances to quantitatively monitor their acquisition and use of the statistics of the stimuli (perceptual prior). We apply the same techniques in the current study to compare the perceptual priors acquired by patients with schizophrenia to those of controls.

Materials & Methods

Participants

A sample of 25 (22 male) individuals with psychosis (diagnosed with either DSM-IV schizophrenia, n = 21; or schizoaffective disorder, n = 4), and 23 (13 male) controls with normal or corrected-to-normal vision were recruited. Patients were recruited from inpatient and outpatient adult mental health services across NHS Lothian. Diagnoses were determined using the Structured Clinical Interview for DSM-IV (SCID-I; First, Gibbons, Spitzer, & Williams, 2002). None of the control participants met DSM-IV criteria for a psychotic disorder, bipolar disorder, or schizotypal or schizoid personality disorder. Symptom severity was measured with the Positive and Negative Syndrome Scale (PANSS) (Kay, Fiszbein, & Opler, 1987), current IQ with the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) and pre-morbid IQ with the National Adult Reading Test (NART; Nelson, & Willison, 1991). All patients were medicated (85% on second generation anti-psychotics, 50% of these were also on mood stabilisers). The study was conducted in accordance with the national and international

ethical standards for human experimentation and research (Declaration of Helsinki, 2013; Good Clinical Practice, 2014). All participants provided fully informed written consent. The study received ethical approval from the South East Scotland Research Ethics Committee 01 and NHS Lothian Research & Development. Following previous studies using the same paradigm (Chalk *et al.* 2010), we determined that 20 participants per group would give us ≥80% power to detect the correct acquisition of the prior and significance between groups (see power calculations in supplemental material). We thus aimed to recruit between 20-25 participant per group to account for possible exclusions due to poor performance at the task.

Twenty patients and twenty three controls successfully performed the task (see below and **Supplementary Figure 1** for exclusion criteria based on performance). The demographic details of the included participants are shown in **Table 1**. It is of note that the patients all had relatively low levels of symptoms on the PANSS, but ongoing functional impairment.

Characteristics	Controls (n=23)	Patients with schizophrenia (n=20)	Significance level p-value
Gender (males)	13	17	0.04
Age (years)	33.86 (12.08)	39.40 (9.43)	0.07
Premorbid IQ	115.55 (4.41)	113.18 (8.76)	0.58
Current IQ	117.45 (7.28)	111.15 (10.70)	0.06
PANSS Positive Scale	8.86 (2.29)	12.55 (5.01)	<0.01
PANSSNegative Scale	8.41 (2.24)	12.20 (5.05)	<0.01
PANSS General Scale	22.27 (6.48)	26.10 (8.35)	0.12
PANSS Total	39.55 (9.22)	50.85 (15.86)	<0.01
GAF	74.81 (11.42)	54.75 (14.36)	<0.001
OLZ eq. (mg/day)	---	12.61 (6.24)	N/A
Illness duration (years)	---	13.33 (9.01)	N/A

Table 1. Participants' demographics. PANSS = Positive and Negative Symptom Scale (lower score is better): potential range is 7-49 for the Negative and Positive scales is, 16-112 for the General scale and 30-210 for the Total score. GAF = Global Assessment of Functioning (potential range 1-100; higher score is better), OLZ eq. = Olanzapine equivalent dosage in mg/day. Values indicate mean and (standard deviation). For gender, group comparisons were done using Chi-square test, for all other measures, two-sided Wilcoxon rank-sum test was used. See Table S1 and Supplementary Figure 6 for a more detailed description of the clinical characteristics.

Apparatus, Stimuli & Procedure

The setup for this study was similar to that used by Chalk *et al.* (2010). Motion stimuli consisted of a field of dots with a density of 2 dots/deg^2, moving coherently (100%) at a speed of 9°/s.

Each trial was composed of two tasks arranged as follows (**Fig. 1A**): First, participants were presented with a fixation point (0.5° diameter) for 400 ms. With the fixation point still on-screen, the motion stimulus (field of dots) was displayed along with a red bar extending from this fixation point. During the presentation of the field of dots, participants were required to estimate the direction of motion by aligning the red bar into the perceived direction of motion (*Estimation task*). The angle of this bar was randomized on each trial and participants were instructed to focus their gaze on the fixation point throughout the estimation task. The display then cleared when either the participant clicked the mouse to validate their choice (estimation) or 3000 ms had elapsed. After the estimation, a 200 ms delay was enforced before the detection screen was presented. The new screen was divided in two equal areas reading 'Dots' and 'No Dots', giving the participants a two-alternative forced choice (2-AFC). Participants were required to move the cursor to the right or the left to indicate whether they detected dots or not, and click to validate their choice (*Detection task*). The cursor then flashed green or red for correct or incorrect responses respectively. No time-outs were enforced during the detection task. Finally, the screen was cleared for 400 ms before a new trial began. Every 20 trials, participants were presented with feedback on their estimation performance in terms of average estimation error (in degrees).

Design

Participants completed 567 trials (i.e. lasting approximately 40 minutes *vs.* 850 trials lasting 60 minutes in Chalk *et al.* 2010) with opportunities for breaks every 170 trials to prevent fatigue. Stimuli were presented at four different randomly interleaved contrast levels. The highest contrast level was 1.7 cd/m^2 above the 5.2 cd/m^2 background. There were 167 trials at zero

contrast and 67 trials at high contrast. Contrasts of other stimuli were determined using a 4/1 and 2/1 staircase on detection performance (García-Pérez, 1998). Throughout the experiment, there were 90 trials with the 2/1 staircase and 243 trials with the 4/1 staircase. On a given trial, the direction of motion for the two staircased contrast levels could either be 0°, ±16°, ±32°, ±48°, and ±64° with respect to a central reference angle. This central reference angle was randomised for each participant.

Unbeknownst to participants, we manipulated their expectations about which motion directions were most likely to occur by presenting stimuli moving at ±32° more frequently (resulting in a bimodal distribution, **Fig. 1B**). At the highest contrast level, 50% of trials were at ±32° and 50% remaining trials at random directions (i.e. not just the predetermined directions).

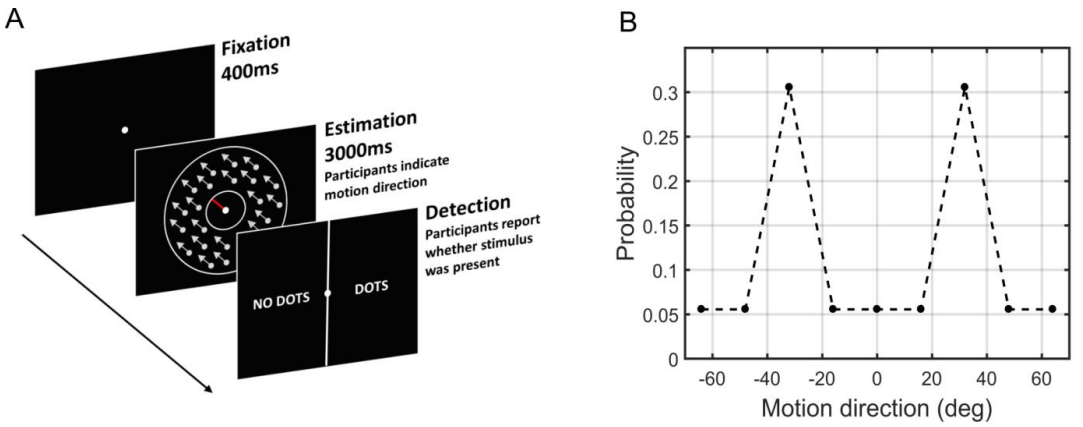


Figure 1: (A) Experimental procedure. Participants were presented with a fixation point followed by the motion stimulus and a response bar (red bar) that they were instructed to align to the perceived motion-direction. The screen was cleared either when participants clicked to validate their estimation or 3000 ms had elapsed. A new screen appeared with a two-alternative forced choice task (2-AFC), requiring participants to indicate whether they perceived the dots during the estimation task. **(B)** Probability distribution of the motion directions. Unbeknownst to participants, the distribution of motion direction was bimodal (i.e. stimuli appeared most often at ±32° from a central direction). The central direction was randomised for each participant.

Behavioural data analysis

Performance on high contrast trials was used as an indicator of whether participants were performing the task adequately. Detection accuracy of at least 70% and estimation root mean square error (RMSE) of less than 30 degrees were the minimum criteria. All 23 controls met these criteria, while 5 out of 25 patients did not meet at least one of the criteria and thus were excluded from further data analysis (**Supplementary Fig. 1**).

The main data analysis was performed on 2/1 and 4/1 staircased contrast levels and only on confirmed trials (i.e. trials where participants validated their choice with a click within 3000 ms in the estimation task and reported seeing dots in the subsequent detection task). The first 100 trials were excluded from the analysis to allow the staircases to converge to stable contrast levels (**Supplementary Fig. 2**). After removing these trials, the luminance levels achieved by the 2/1 and 4/1 staircases were found to be overlapping (**Supplementary Fig. 2**), thus they were combined for all further analysis. Finally, since the distribution of presented directions was symmetrical around a central reference angle, the behavioural measures at equal absolute distance from the reference angle were averaged together.

In the estimation task, the variance of participants' direction estimates was large. As in previous work (Chalk *et al.*, 2010; Gekas *et al.*, 2013), we hypothesized that this variability resulted from random estimations on a proportion of trials, thus increasing substantially the variance of motion-direction estimates. To account for this, we fitted the individual estimation responses to the following distribution:

$$(1-\alpha)\cdot V(\mu,\sigma) + {}^{\alpha}/_{2\pi} \quad (1)$$

where α is the proportion of estimations at random directions and $V(\mu, \sigma)$ is the circular normal (i.e. von Mises; Mardia, 1972) distribution with mean μ and width σ.

In trials where no stimulus was presented, we reconstructed the probability distributions of participants' responses over motion directions using Kernel Density Estimation (Silverman, 1986; Wand, 1994). The KDE is a non-parametric method used to estimate the probability

density function from discrete measures of a continuous variable. To do so, a kernel that defines the form of the probability density function (e.g. von Mises kernel) was placed at each of the observed measurement. Then, all the individual kernels were summed to create the probability density function of the random variable (motion direction).

To assess the effects of the acquired expectations in both groups and to simultaneously determine whether patients differed from controls, we performed 2-way ANOVA with motion direction as a within-subjects factor and group as a between-subjects factor; and Bonferroni post hoc tests for all pairwise comparisons. All analysis was done using SPSS version 24. To determine the strength of the evidence for the null hypothesis (i.e. finding no group difference vs. finding evidence that the groups are similar) we also report Bayes factors using the Bayesian statistical software package JASP. We report Kendall correlation coefficients whenever the data is ordinal with rank ties and/or strong outliers.

Modelling

To test for individual variability in the underlying perceptual inference and to obtain more direct measurements of the acquired expectations, we fitted a range of models to our data. The first class of models assumed that the biases were of a perceptual nature, as conceived in the Bayesian framework: sensory information is combined with a learned prior of the stimulus statistics in a probabilistic way. The simple 'BAYES' model assumed that the likelihood precision was constrained to be the same across all presented motion directions (corresponding to the hypothesis that there was no learning in the likelihood due to the distribution of the motion directions). An additional variant of the 'BAYES' model tested the hypothesis that lapse estimations were not completely random, but instead were made according to the acquired prior expectations. We call these responses 'prior-based lapses' (**Fig. 4**). This model was termed 'BAYES_P' and was otherwise equivalent to 'BAYES'.

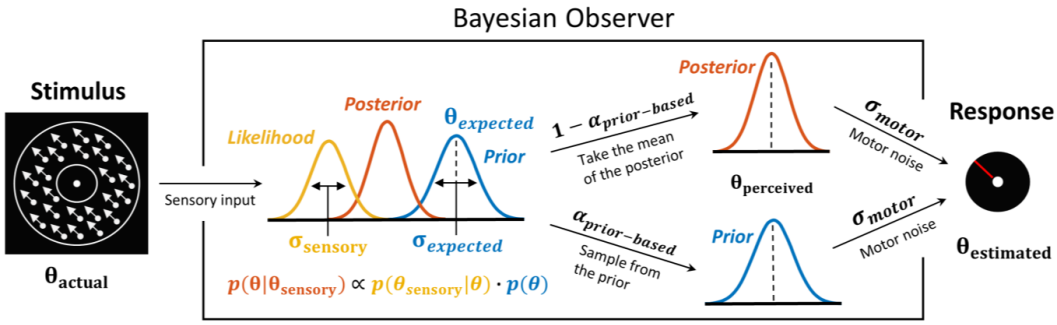


Figure 4. Bayesian model of estimation response for a single trial. The actual motion direction (θ_{actual}) is corrupted by sensory uncertainty ($\sigma_{sensory}$), and then combined with prior expectations (mean $\theta_{expected}$ and uncertainty $\sigma_{expected}$) to form a posterior distribution. The perceived motion direction ($\theta_{perceived}$) then corresponds to the mean of the posterior distribution. However, on a fraction of trials, determined by the lapse rate ($\alpha_{prior\text{-}based}$), the perceived motion direction is sampled form the prior. Finally, in both cases, the response ($\theta_{estimated}$) is made by perturbing $\theta_{perceived}$ with motor noise (σ_{motor}). This results in 4 free model parameters: $\sigma_{sensory}$, $\sigma_{expected}$, $\theta_{expected}$ and $\alpha_{prior\text{-}based}$. The motor noise ($\sigma_{motor}$) is estimated from high contrast trials and is used as a fixed parameter.

Another class of models assumed that task performance could be explained by response strategies that do not involve Bayesian integration (heuristic models). According to these models, on any given trial participants responses were based solely on either the prior expectations or sensory information. We considered four variations of response strategy models (see Methods and Supplementary information for details). Below we present the Bayesian models as they provided a better explanation to the data (see **Fig. 5**, model comparison).

Bayesian model

Following the Bayesian framework, we assumed that participants combined sensory information (likelihood) with their expectations about the motion direction (prior) on every trial. The sensory likelihood of the observed motion direction ($\theta_{sensory}$) was parameterised as a von Mises circular normal distribution with variance $\sigma_{sensory}$:

$$p_{likelihood}(\theta_{sensory}|\theta) = V(\theta, \sigma_{sensory}) \quad (2)$$

The mean of this distribution depended on the actual presented motion direction (θ_{actual}), and to account for trial-to-trial variability it was drawn from another von Mises distribution centred on θ_{actual} with variance $\sigma_{sensory}$ (i.e. $V(\theta_{actual}, \sigma_{sensory})$).

We then hypothesized that participants acquire an approximation of the 'true' prior from experience (p_{prior}), representing the participants' expectations of motion directions. The acquired priors were parameterized as the sum of two von Mises circular normal distributions, centred on motion directions $\theta_{expected}$ and $-\theta_{expected}$, each with variance $\sigma_{expected}$:

$$p_{prior}(\theta) = \frac{1}{2}[V(-\theta_{expected}, \sigma_{expected}) + V(\theta_{expected}, \sigma_{expected})] \quad (3)$$

Combining the prior and the likelihood gives us the posterior probability that the stimulus is moving in a direction θ:

$$p_{posterior}(\theta|\theta_{sensory}) \propto p_{likelihood}(\theta_{sensory}|\theta) \cdot p_{prior}(\theta) \quad (4)$$

The perceived direction, $\theta_{perceived}$, was taken to be the mean of the posterior distribution (almost identical results would be obtained by using the maximum instead).

Finally, we accounted for motor noise (i.e. aligning and clicking the mouse) and lapse estimations, such that:

$$p(\theta_{estimate}|\theta_{perceived}) = (1 - \alpha_{prior\text{-}based}) \cdot V(\theta_{perceived}, \sigma_{motor}) + \alpha_{prior\text{-}based} \cdot p_{prior}(\theta) \quad (5)$$

where σ_{motor} is the motor noise and $\alpha_{prior\text{-}based}$ is the proportion of prior-based lapse estimations on each trial (i.e. lapse estimations that follow the participants' acquired expectations – $p_{prior}(\theta)$). That is we posited that when participants are shown a stimulus below their contrast threshold, it effectively becomes a no-stimulus trial, where participants might 'hallucinate' a stimulus at the most expected motion directions, sampling from their prior distribution

(Laquitaine and Gardner 2018, see **Fig. 7**). We called this model 'Bayes_P' for Bayes with Prior-based lapses.

Finally, we also tested variants of this model, where lapse estimations were uniformly distributed, rather than following the participants expectations (model 'BAYES'), or that due to increased exposure to stimuli at specific angles, sensory uncertainty $\sigma_{sensory}$ could vary across angles (0°, ±16°, ±32°, ±48°, ±64°; model 'BAYES_var'), or that sensory uncertainty varied only at the most presented directions (model 'BAYES_varmin' – see Methods and Supplementary information for details).

Results

Detection performances and contrast levels

Participants' detection performance was monitored to adapt the stimulus contrast to each participant's just noticeable difference (JND, a measure of contrast sensitivity). Using 2/1 and 4/1 staircases, we ensured that the individual detection performances would converge to 70.4% and 84.1% respectively (Levitt, 1971).

Contrast staircases converged to stable luminance levels after about 100 trials for both groups (**Supplementary Fig. 2**); Controls converged to 0.41 cd/m^2 (±0.03) for 2/1 staircase and 0.46 cd/m^2 (±0.03) for 2/1 staircase while patients converged to 0.57 cd/m^2 (±0.04) for 2/1 staircase and 0.62 cd/m^2 (±0.05) for 4/1 staircase. These results confirm previous findings (Skottun and Skoyles, 2007) suggesting that patients with schizophrenia display significantly poorer contrast-sensitivity in comparison to controls (2/1 staircase: $Z = 3.15$, $p = 0.002$; 4/1 staircase: $Z = 2.90$, $p = 0.004$, two-tailed rank-sum test).

Statistical learning

First, we investigated whether participants acquired the statistics of the stimulus. To do so, we looked at patterns suggestive of statistical learning in each group, namely attractive biases

towards the most frequent directions, decreased reaction times and improved detection performance for the most frequent directions (**Fig. 2**).

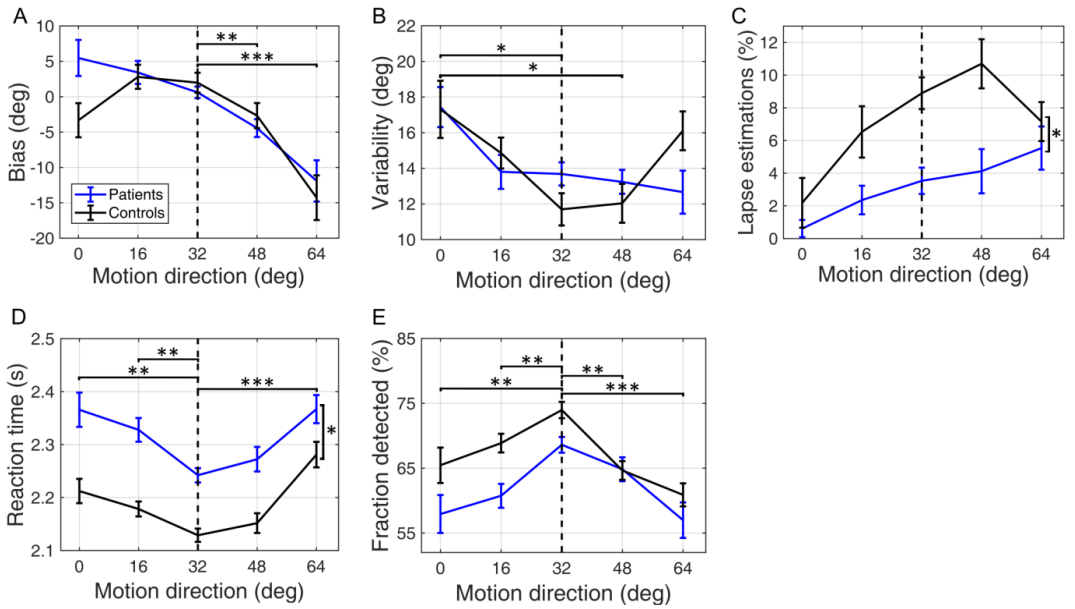


Figure 2: Performance on low contrast trials by patients (blue lines) and controls (black lines). **(A)** Mean estimation bias as a function of the true motion direction. **(B)** Estimation standard deviation (i.e. variability) as a function of the presented of motion direction. **(C)** Lapse estimations as a function of motion direction, estimated using (eq. 1). **(D)** Reaction times during the estimation task as a function of motion direction. **(E)** The fraction of trials in which the stimulus was detected as a function of the presented motion direction. The error bars represent within-subject standard error. The vertical dashed lines correspond to the most frequently presented motion directions (i.e. ±32°). *, ** and *** indicate significance levels at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. These results are not affected by any violations of the normality assumption; see 'Normality and Non-parametric testing' in the supplementary information.

Estimation performance

To investigate whether the participants' perceived motion-directions were biased, we measured the difference between the true motion direction and the motion direction reported by the

participants. **Fig. 2A** displays the average estimation bias plotted against the true motion direction for each group. Overall, there was a significant effect of motion direction on the estimation bias ($F(2.45, 100.52) = 15.37$, $p < 0.001$, $\eta_p^2 = 0.273$, Greenhouse-Geisser correction $\varepsilon = 0.613$), but no differences between the groups (group main effect: $F(1, 41) = 0.83$, $p = 0.369$, $\eta_p^2 = 0.001$; with moderate to substantial evidence for the null hypothesis, $BF_{01} = 3.99$); and no group*angle interaction ($F(2.45, 100.52) = 1.64$, $p = 0.193$, $\eta_p^2 = 0.038$). Pairwise comparisons (with Bonferroni correction) revealed that there was an attractive bias towards ±32° at ±48° and ±64° (MD = 4.858, p = 0.002; and MD = 14.395, p < 0.001, respectively), but not at ±16° (MD = 1.818, p = 0.955). Together, these results confirm that both patients and controls were biased towards perceiving motion directions as being more similar to the most frequently presented directions than they really were, consistent with having acquired the priors that approximate the statistics of the stimulus.

We also investigated whether the effects of acquired prior expectations were reflected in the variability of estimations, namely a decrease of variability for the expected directions (**Fig. 2B**). We found a significant main effect of motion direction ($F(3.07, 125.69) = 5.18$, $p = 0.002$, $\eta_p^2 = 0.112$, Greenhouse-Geisser correction $\varepsilon = 0.766$), but no differences between the groups (main effect of group: $F(1, 41) = 0.02$, $p = 0.880$, $\eta_p^2 = 0.001$; with moderate to substantial evidence for the null hypothesis, $BF_{01} = 3.74$); and no group*angle interaction $F(3.07, 125.69) = 1.58$, $p = 0.196$, $\eta_p^2 = 0.037$). Pairwise comparisons showed that variability at 0° stood out the most, being significantly larger than at ±32° and ±48° (MD = 4.680, p = 0.012; and MD = 4.733, p = 0.025, respectively), although not different than at ±16° and ±64° (MD = 3.044, p = 0.239 and MD = 2.990, p = 0.541). The increased variability in the region between the two modes reflects their conflicting influence on the percepts in this region.

Finally, we analysed lapse estimations, which were captured by the 'α' term in Eq. (1) and which were assumed to arise from random responses on some of the trials (**Fig. 2C**). We found both motion direction and group main effects to be significant ($F(4,164) = 5.76$, $p < 0.001$, $\eta_p^2 = 0.123$ and $F(1, 41) = 6.41$, $p = 0.015$, $\eta_p^2 = 0.135$, respectively), with patients exhibiting fewer lapse estimations. Pairwise comparisons revealed that lapse rate at 0° was significantly smaller than at all other directions (±32°, MD = 4.814, p = 0.001; ±48°, MD = 6.010, p = 0.007; ±64°,

MD = 4.951, p = 0.003), except for ±16° (MD = 3.043, p = 0.393). The finding that the estimated lapses would depend on the presented motion direction was surprising: if the lapses were completely random, we should observe no such dependence in the estimated values. We therefore postulated that when participants make guesses about the direction of motions, their estimations might be sampled from the distribution of the acquired prior expectations rather than uniformly. Later, we explicitly tested this possibility via computational modelling and found that it indeed was the case (see Modelling results).

Reaction times and Detection Performance

Next, we examined how participants' acquired expectations influenced reaction times and the detection of stimulus. The estimation reaction times (**Fig. 2D**) show a significant main effect of motion direction ($F(2.73,111.76) = 10.80$, $p < 0.001$ $\eta_p^2 = 0.209$, Greenhouse-Geisser correction $\varepsilon = 0.681$). This was driven by decreased reaction times at the most frequent directions as revealed by pairwise comparisons: reaction time at ±32° was significantly shorter than at all other directions (0°, MD = 0.104, p = 0.001; ±16°, MD = 0.068, p = 0.004; ±64°, MD = 0.139, p < 0.001), except for ±48° (MD = 0.027, p = 1.000). Furthermore, patients were found to also be significantly slower than controls ($F(1, 41) = 4.11$, $p = 0.049$, $\eta_p^2 = 0.091$), but we found no interaction between group and motion direction ($F(2.73, 111.76) = 0.66$, $p = 0.563$, $\eta_p^2 = 0.016$). Slow reaction time is a hallmark of schizophrenia that has been documented thoroughly in the literature in simple reaction-time tasks using visual and/or auditory stimuli (e.g., see Nuechterlein, 1977; Fioravanti *et al.*, 2012).

An even more direct way of assessing how the acquired expectations influenced the detection of stimulus is to analyse the fraction of trials where participants explicitly report seeing or not seeing the stimulus (**Fig. 2E**). We found that the detection of stimulus was greatly affected by the presented motion direction ($F(2.36,96.64) = 8.51$, $p < 0.001$, $\eta_p^2 = 0.172$, Greenhouse-Geisser correction $\varepsilon = 0.589$), with stimulus at ±32° being the most frequently detected direction as shown by pairwise comparisons: detection at ±32° was significantly better than at all other directions (0°, MD = 9.59, p = 0.004; ±16°, MD = 6.48, p = 0.001; ±48°, MD = 6.54,

p = 0.001; ±64°, MD = 12.35, p < 0.001). However, the groups were not found to be different (main group effect: F(1,41) = 3.62, p = 0.064, η_p^2 = 0.081; although there was no evidence for the null hypothesis either, BF_{01} = 0.97); group*motion direction interaction was also non-significant: F(2.36,96.64) = 1.11, p = 0.340, η_p^2 = 0.026).

Overall, these results indicate that, in terms of detection responses (hit rates and reaction-time), similar benefits of statistical learning were present in both patient and control groups. Overall, behavioural measures suggest that prior effects (e.g. Bias, RT, and Hit rate) became significant as early as within first ~100 trials for both patients and controls (see **Supplementary Figures 3-4**).

Perceived motion in absence of visual stimuli (hallucinations)

Finally, we investigated whether the acquired statistics about the motion stimulus affected the participants' perception on trials where no stimulus was presented, but where participants reported both a motion direction and seeing a stimulus. We refer to this effect as hallucinations. These hallucinations in our perceptual task are of course different in terms of content and complexity from the visual hallucinations observed in psychosis. However, the underlying mechanisms might be informative to the understanding of illusions and hallucinations in schizophrenia (Silverstein and Keane 2011a; 2011b; Notredame *et al.*, 2014).

To quantify the probability ratio that participants made estimates that were closer to the most frequently presented motion directions relative to other directions, we multiplied the probability that participants estimated within 16° of these motion-directions by the total number of 32° bins:

$$p_{ratio} = p(\theta_{estimate} = \pm 32(\pm 16)°) \cdot N_{bins} \qquad (6)$$

This probability would be equal to 1 if participants were equally likely to estimate within 16° of ±32° as they are to estimate within the other 16° bins.

We found that the median value of *'p_{ratio}'* was significantly greater than 1 for both patients and controls (median(p_{ratio}) = 2.88, p = 0.003 and median(p_{ratio}) = 2.75, p <0.001, respectively; two-tailed signed-rank test), indicating that both patients and controls were much more likely to hallucinate the most frequent motion directions as opposed to all other directions (**Fig. 3A, B**). Bayesian statistical analysis provided moderate to substantial evidence for the groups being the same in this measure (BF_{01} = 3.32). Finally, hallucinations of the most frequent directions (i.e. hallucinations at ±32°±16°) were quick to develop during the course of the experiment: they became significant after only 150 trials for both controls (*p=0.036,* one-tailed signed-rank test; **Supplementary Fig. 3D**) and patients (*p=0.035,* one-tailed signed-rank test; **Supplementary Fig. 4D**).

While both patients and controls hallucinated predominantly towards the most frequently presented directions (i.e. prior-based hallucinations), patients exhibited fewer of such hallucinations (**Fig. 3C**; hallucinations within ±16º of ±32º; p = 0.016, two-sided rank-sum test), and also exhibited less hallucinations overall (**Fig. 3D**; p = 0.004, two-sided rank-sum test). We wanted to know whether the severity of the symptoms was predictive of the magnitude of this effect. However, we found no correlations between the number of hallucinations and the PANSS positive, negative, general or total scores nor duration of illness, or between the daily-dosage of anti-psychotics (Olanzapine equivalent; Leucht et al., 2015) and the total number of hallucinations.

We also investigated the distribution of responses when participants estimated the direction of motion but reported not seeing any dots. Interestingly, in this subset of trials, and unlike in our previous work (Chalk et al, 2010), estimations were also more likely than chance to be made around the most frequent motion directions by both patients and controls (median(p_{ratio}) = 1.24, p = 0.002 and median(p_{ratio}) = 1.41, p = 0.045, respectively; two-sided signed rank test). One explanation for this might be habitual effects - when the bar is moved towards the most frequent directions out of motor habit. Another possible explanation is that participants might have hallucinated stimuli in the expected prior directions, but their confidence about their percept being very low, they sometimes chose to report not seeing any dots in the hope to give the correct answer (each detection response was followed by immediate feedback).

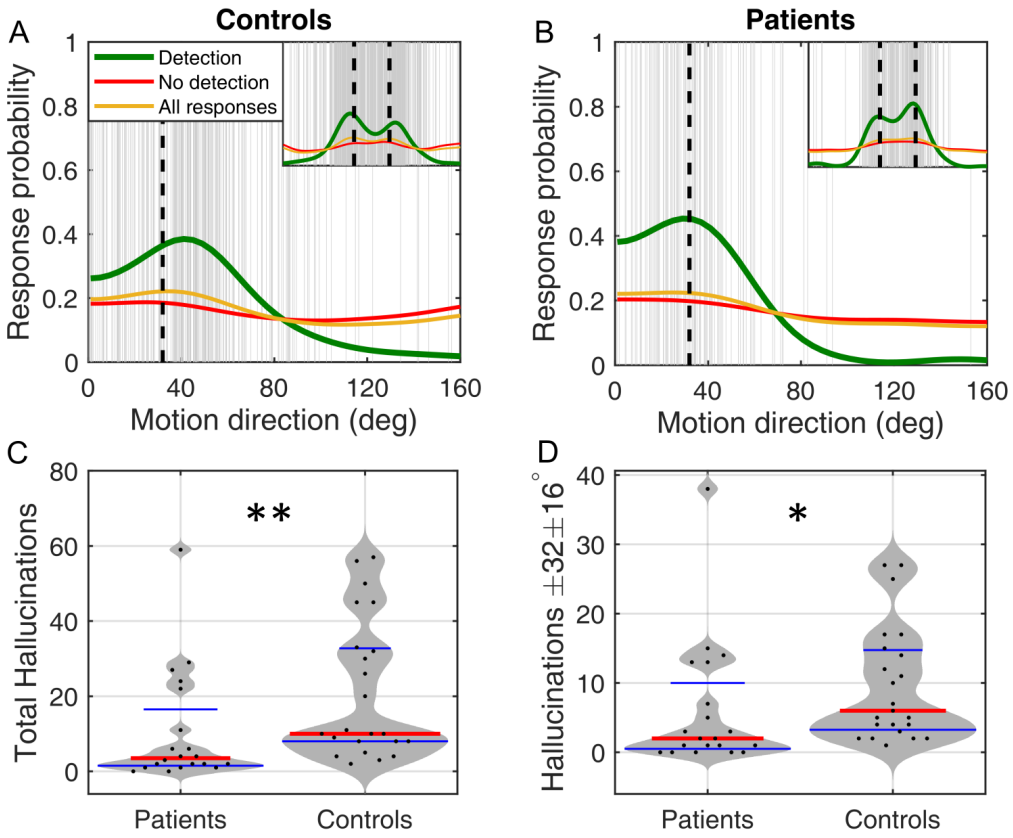


Figure 3. Estimation responses in the absence of stimulus. **(A, B)** Distribution of the estimation responses by patients and controls, respectively. The vertical grey lines represent reported motion directions when no stimulus was present (i.e. hallucinations) pooled across each group. The green line denotes the probability distribution of these hallucinations, which was produced using Kernel Density Estimation. The red line denotes probability distribution of responses that were followed by participants' reporting seeing no stimulus. The orange line denotes all estimations regardless of the detection response. In the main plots the data is averaged across the central motion direction, while the insets show the corresponding distributions across the full range. **(C, D)** Comparison of patients and controls by **(C)** the total number of hallucinations ($p = 0.004$, two-sided rank-sum test) and **(D)** the number of hallucinations around the most frequently presented motion directions (within ±16º of ±32º; $p = 0.016$, two-sided rank-sum test). Red horizontal lines denote median values; blue horizontal lines denote 25th and 75th percentiles. Black dots denote individual participants, grey areas represent density of the data points. * and ** indicate significance levels at $p < 0.05$ and $p < 0.01$ respectively.

Modelling results

We fitted the models to the behavioral data and computed the Bayesian Information Criterion (BIC) for each participant, which gives us information regarding model fit while penalizing for extra model complexity (preventing overfitting). We found that the BAYES_P model had the smallest BIC for both patients and controls (**Fig. 5A, and B**) – indicating best performance, with a difference in BIC between the winning model (BAYES_P) and the second best model (BAYES) being larger than 10. This is equivalent to a log Bayes factor larger than 10, and is considered to be very strong or 'decisive' evidence in favor of the winning model (Kaas & Raftery, 1995). We also ran a random effect Bayesian model selection analysis (Rigoux et al., 2014; Daunizeau et al., 2014) to ensure that the favoured model was not being selected due to a subset of participants in each group. The analysis confirmed that BAYES_P was best at describing behaviour for both groups (**Fig. 5C and D**). Model fits to the data also showed that BAYES_P to be better at fitting lapse estimations, confirming that such estimations followed the acquired prior distribution instead of being random (**Fig. 6C,G**). Moreover, we found a strong correlation between the prior-based lapses recovered via BAYES_P and hallucinations ($\tau_b = 0.657$, $p < 0.001$; Kendall's correlation; **Fig. 7B**), suggesting that prior-based lapses could be considered as hallucinations experienced on trials when stimulus is too weak to be detected (**Fig. 7A**).

Finally, we compared patients and controls on the basis of BAYES_P parameter estimates (**Fig. 6I-L**). Consistent with the behavioral data analysis, we found no differences in the acquired prior expectations (**Fig. 6I, J;** the mean of acquired prior: $p = 0.874$, $BF_{01} = 3.32$; and the uncertainty in the acquired prior: $p = 0.401$; two-tailed rank-sum test; $BF_{01} = 2.95$). There were no differences in the precision of sensory likelihood (**Fig. 6K**, $p = 0.742$, two-tailed rank-sum test; $BF_{01} = 2.96$). Lastly, just as in the behavioral data, we found that patients made less prior-based lapse estimations (**Fig. 6L**, $\alpha_{\text{prior-based}}$: $p = 0.024$, two-tailed rank-sum test), suggesting that the model parameter '$\alpha_{\text{prior-based}}$' captures the participants' propensity to experience prior-based hallucinations in our task.

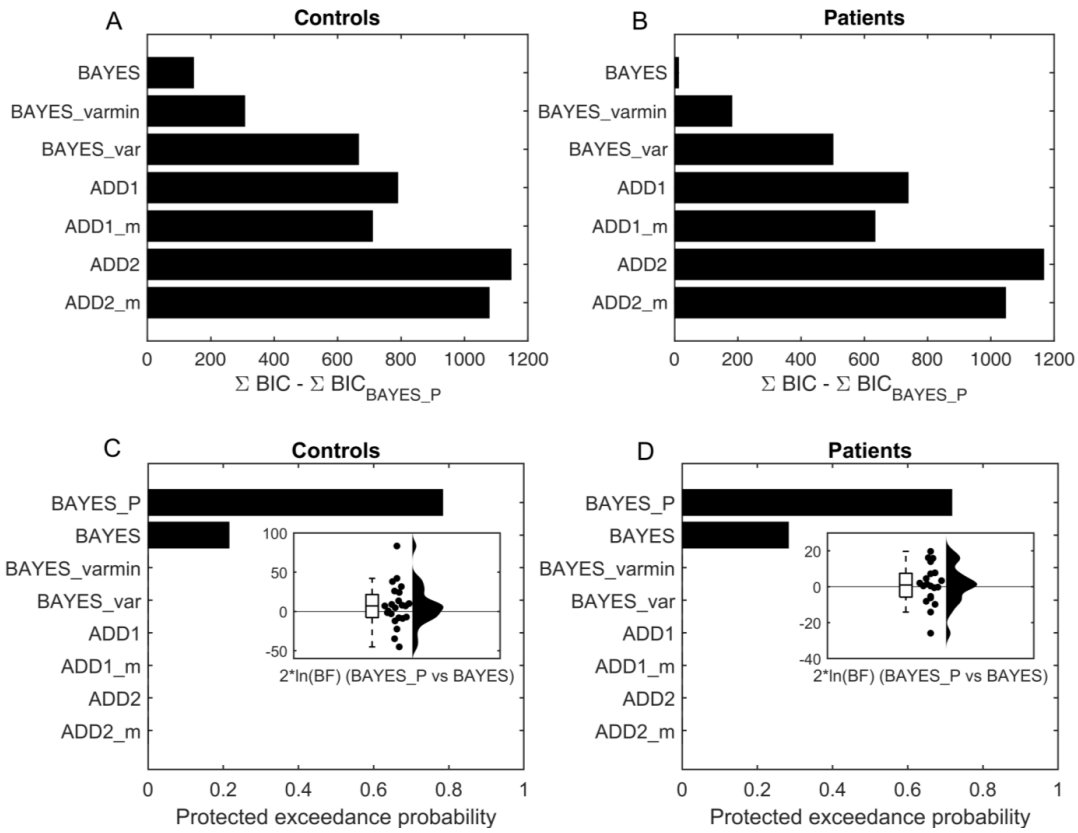


Figure 5. Model comparison and selection. (**A, B**) Fixed effects model selection using Bayesian Information Criterion (BIC) for (**A**) controls and (**B**) patients. X-axis measures the relative difference between BIC of each model (as indicated on Y-axis) and BIC of BAYES_P (winning model) summed across participants. Smaller BIC indicate a better model fit while penalizing for added model complexity. For both patients and controls BAYES_P provided the best model evidence, 12 BIC units better than BAYES for patients and 146 BIC than BAYES for controls. (**C, D**) Random effect Bayesian model selection (Rigoux et al., 2014; Daunizeau et al., 2014) for (**C**) controls and (**D**) patients. Higher protected exceedance probability indicates a model having a higher likelihood of being more frequent among the subjects. For both patients and controls BAYES_P was the most likely model. The insets in **C** and **D** show the distribution of Bayesian Factor for BAYES_P vs. BAYES summarized by a boxplot, jittered data scatter and a probability density which was estimated using a normal kernel.

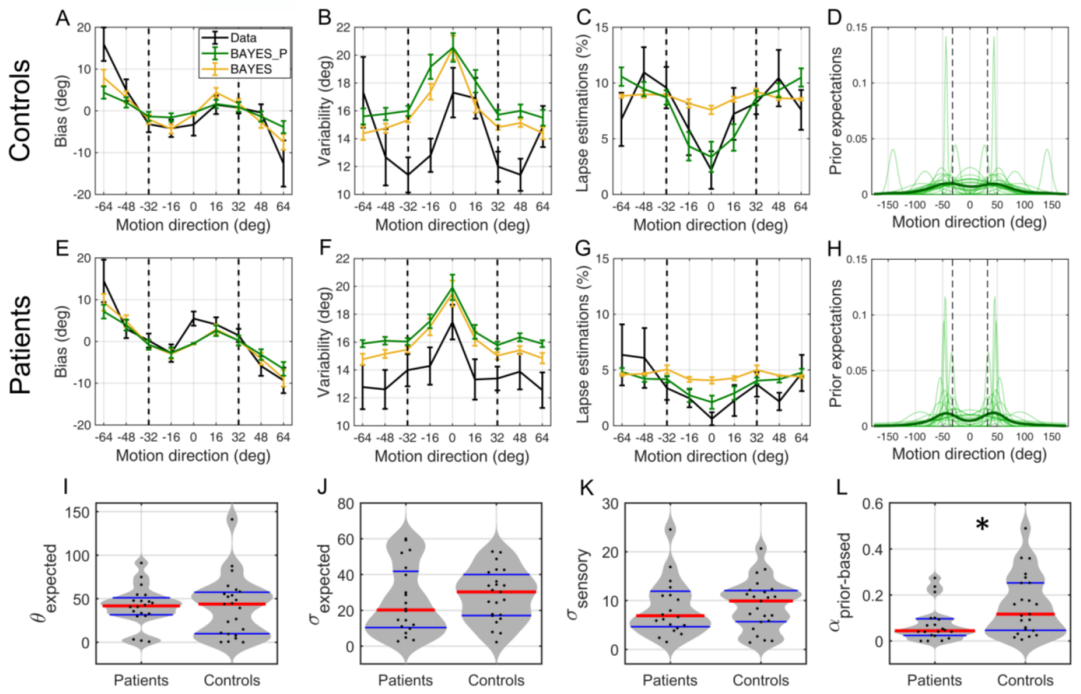


Figure 6. Model fits and parameter estimates. **(A-H)** Model fits for the best fitting model BAYES_P (green) and the second best model BAYES (yellow), to the behavioral data (black). **(A-D)** controls and **(E-H)** patients. **(A, E)** estimation bias, **(B, F)** estimation variability, **(C, G)** estimation lapse rate, **(D, H)** prior expectations of each individual (transparent green) and group average (thick green) as estimated via BAYES_P model. The vertical dashed lines correspond to the most frequently presented motion directions (i.e. ±32°). The error bars represent within-subject standard error. **(I-L)** Comparison of BAYES_P model parameter estimates of patients and controls. **(I)** $\theta_{expected}$ – the mean of acquired prior ($p = 0.874$, two-tailed rank-sum test; BF01 = 3.32), **(J)** $\sigma_{expected}$ – the uncertainty in the acquired prior ($p = 0.401$, two-tailed rank-sum test; BF01 = 2.95), **(K)** $\sigma_{sensory}$ – the uncertainty of sensory likelihood ($p = 0.742$, two-tailed rank-sum test; BF01 = 2.96), **(L)** $\alpha_{prior\text{-}based}$ – prior-based lapse rate ($p = 0.024$, two-tailed rank-sum test). Red horizontal lines denote median values; blue horizontal lines denote 25th and 75th percentiles. Black dots denote individual participants, grey areas represent density of the data points. * indicates significance level at $p < 0.05$. See Supplementary Figures 7 and 8 for the model fits for bias of each individual.

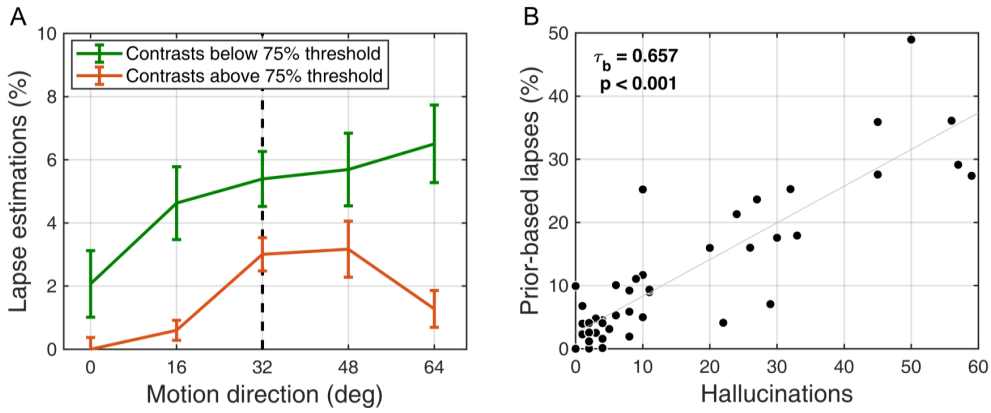


Figure 7. Relationship between lapse estimations and hallucinations. (**A**) The amount of lapse estimations at different stimulus contrast levels. Contrast staircase trials were split into two subsets along the 75% detection threshold, which was determined for each individual from their psychometric curves. The data then was pooled across both patient and control groups. We found that the amount of lapse estimations depended on the presented contrast level with more lapse estimations being made at lower contrasts ($F(1,84) = 12.61$, $p < 0.001$), which supported our interpretation that these estimations were not merely attentional lapses but hallucinations on trials when stimulus was hard to detect. The error bars represent within-subject errors. (**B**) Prior-based lapses and hallucinations. A strong positive correlation between prior-based lapse estimations (recovered via BAYES_P model) on low contrast trials and hallucinations on no-stimulus trials ($\tau_b = 0.657$, $p < 0.001$; Kendall's correlation) provided further support that both of these behaviours are driven by the same mechanism (i.e. both being samples from the prior).

Parameter recovery for model BAYES_P

To gauge the reliability of our modelling results, we performed parameter recovery for the winning BAYES_P model. Parameter recovery consists of simulating synthetic data with different sets of known parameter values ('true parameters') for a given model and then fitting the same model to the synthetic data to estimate and recover these parameters ('recovered parameters'). The strength of correlation between the actual and recovered parameters measures the reliability of modelling results. Parameter recovery, just as the parameter estimation from behavioral data, is sensitive to any correlations that might be present among

the model parameters, to the choice of parameter estimation methods and also to the amount of data used for model fitting. Therefore, parameter recovery serves as a crucial step in validating the reliability of the modelling results (Palminteri, et al., 2017).

We found that the winning BAYES_P model recovered parameters very well, which was reflected in the coefficient of determination (R^2) for all recovered parameters being $R^2 \geq 0.84$ (**Supplementary Fig. 9**).

Discussion

We were interested in testing the emerging model of schizophrenia proposing that the disorder could stem from deficits in Bayesian inference (Corlett *et al.*, 2009a, 2009b; Fletcher and Frith, 2009; Adams *et al.*, 2013; Schmack et al., 2013, 2015, 2017; Teufel et al., 2015; Powers et al., 2017, Jardri et al., 2017). The experimental paradigm we chose is well suited to quantitatively assess the acquisition of sensory priors, how these priors are used in perception, as well as to quantify inter-individual variability in the learning and inference process (Chalk et al 2010, Karvelis et al, 2018).

Acquisition of visual prior expectations

We found that both the control and patient groups implicitly learned the statistics of the motion stimuli and that those expectations modified their perception, consistent with them acquiring a Bayesian prior of the stimulus statistics and combining it with sensory evidence, replicating our previous results (Chalk et al, 2010). This was reflected by attractive estimation biases towards the frequently presented directions, faster reaction times and higher detection rates at these directions, as well as hallucinated motion directions in the absence of stimulus predominantly following the most frequent directions.

Patients with schizophrenia were not qualitatively, nor quantitatively different from controls in the measures used to assess learning of the task statistics. This suggest that while there are clearly domain-specific perceptual (e.g. hallucinations) and learning deficits in schizophrenia (e.g. jumping to conclusions), our study demonstrates that these deficits are not due to a domain-general impairment in the acquisition and/or utilization of statistical information in the environment. That is, we find that patients with chronic schizophrenia do not appear to be impaired in the acquisition of visual statistical priors in our task.

These results are consistent with studies finding no deficit in implicit learning in schizophrenia (Kéri *et al*., 2000; Danion *et al*., 2001; Marvel *et al*., 2005; for review see: Gold *et al*., 2009). In contrast with studies that assay explicit statistical learning and inference using more cognitive tasks (i.e. usually believed to involve frontal cortical regions), here we measured implicit statistical learning of visual stimuli that could be embodied in visual processing areas rather than frontal cortices (Kok *et al*., 2013). In fact, patients with schizophrenia appear relatively spared in implicit learning tasks that do not require integrating information after each trial (Gold *et al*., 2009). These results are also consistent with our previous study using the same paradigm showing intact statistical learning in participants with high schizotypal traits (Karvelis et al, 2018).

Impact of acquired visual prior expectations

We found no difference between patients and controls as to the influence of the acquired expectations on their performance regarding estimation of the motion directions. They were not more or less biased towards the most frequent directions, nor more or less variable in those estimations.

Patients were found to differ from controls in three ways, however. First, patients with schizophrenia displayed significantly poorer contrast discrimination thresholds and slower reactions times, as documented in previous studies. Second, and more interestingly, patients reported significantly fewer hallucinations at all directions and fewer prior-based hallucinations (i.e. hallucinations of the most frequently presented motion directions). Third, patients exhibited fewer prior-based lapse estimations than controls on low contrast trials. Prior-based

lapses can be interpreted in the same way as hallucinations on no-stimulus trials. When using contrast staircases, contrast levels hover around the detection threshold, which means that on a significant number of trials the stimulus contrast falls below the participants' threshold of perception, effectively becoming equivalent to trials with no stimulus. If participants have hallucinations on these trials (and thus report that they have perceived a stimulus), these will be expressed in our results as prior-based lapses. In support for this interpretation, we found that there were significantly more lapse estimations made on trials when the stimulus contrast was below the 75% detection threshold then when it was above ($F(1,84) = 12.61$, $p < 0.001$; **Fig. 7A**) and we found a strong correlation between prior-based lapses and hallucinations on no-stimulus trials (**Fig. 7B**). Together, this suggests that although patients appear to acquire the same prior as controls, they tend to hallucinate this prior less than controls when there is no stimulus, or when the stimulus is below detection threshold.

The fact that patients exhibit fewer prior-based hallucinations suggests that their perception is less influenced by their learned prior expectations than observed in control participants. This is consistent with previously reported findings suggesting that patients with chronic schizophrenia are less sensitive to expectation-driven illusions (e.g. the hollow-mask illusion) than controls (Tschacher *et al*., 2006; Dima *et al*., 2009; Crawford *et al.*, 2010; Horton and Silverstein, 2011; Keane *et al.*, 2013, Notredame et al., 2014). This finding is also in line with results from Schmack et al. (2013, 2015), reporting a decreased influence of induced expectations (priors) on perception.

It is intriguing however that the influence of prior expectations is weaker in the detection task, but similar to that of controls for the estimation task. The absence of differences in the estimation bias could be explained by different factors. One factor might be that very few of our patients reported clinically significant levels of hallucinations (PANSS items >3), and that this propensity might be more pertinent than the diagnosis of schizophrenia *per se* (Powers et al, 2017). We found some evidence supporting this idea: PANSS Positive symptom score and prior-based lapses (estimated via BAYES_P) were close to being negatively correlated (τ_b=-0.314, p= 0.063; Kendall's correlation; **Supplementary Fig. 5A**); this relationship was also backed up by a much stronger correlation for when lapse estimations were estimated directly from the behavioural data using Eq. (1) (τ_b = -0.465, p = 0.006; Kendall's correlation;

Supplementary Fig. 5B). However, it should be noted that no correlation was found between the PANSS Positive symptom score and hallucinations exhibited on no-stimulus trials.

Similarly, the absence of stronger effects might be related to illness duration: weaker estimation biases may be characteristic of earlier stages of the illness but may not be detectable anymore in our patients as they have been generally ill for a long time, due to medication or compensatory mechanisms. We find some support for this idea with a significant correlation between duration of illness and magnitude of estimation bias ($\tau_b = 0.523$, $p = 0.003$; Kendall's correlation; **Supplementary Fig. 5C**).

Finally, since the differences appear only for the detection part of the task, it might be that chronic patients have simply developed an increased perceptual threshold. Following this idea, patients would require stronger evidence (i.e. sharper posterior) in order to perceive a stimulus or to make a decision about the presence of a stimulus. This is consistent with the fact that patients required higher stimulus contrasts and integrated information over longer periods of time before responding (slower reaction times during the estimation task). We therefore hypothesize that it is a possible adaptation strategy used by patients over time to minimize responses to stimuli that were not truly present (i.e. their psychotic hallucinations).

Taken together, our results suggest that statistical learning is intact in relatively well patients with chronic schizophrenia on stable doses of second generation antipsychotic medication. The impact of their acquired priors is also the same as that of controls in the estimation task, but weaker in the detection task. These results are surprising in view of the current prominent theories proposing that schizophrenia is a disorder of predictive processing or Bayesian inference, and suggest ways in which these accounts need to be nuanced. The similarity of controls and patients' performance in our task may however be related to the success of treatment or some other adaptive process related to illness duration. Future work will aim at testing participants at earlier stages of the illness, and ideally before pharmacological treatment has begun.

Funding

VV was supported by grants EP/F500386/1 and BB/F529254/1 for the University of Edinburgh School of Informatics Doctoral Training Centre in Neuroinformatics and Computational Neuroscience (www.anc.ed.ac.uk/dtc/) from the UK Engineering and Physical Science Research Council (EPSRC), and the UK Medical Research Council (MRC). PK was funded by Engineering and Physical Sciences Research Council. PS was supported by the NARSAD grant (19271). ARS was funded by NSF Grant BCS-1057625 and NIH grant 1R01EY023582. The collaboration between PS and ARS was supported by a Marie Curie Exchange Grant FP7-PEOPLE-2009-IRSES-247543. Some of this work was completed while PS was visiting the Simons Institute for the theory of Computing. Psychophysical testing, neuropsychological testing and structured clinical interviews were conducted at the University of Edinburgh, Royal Edinburgh Hospital. The funding bodies had no control over the design, analysis or acquisition of the data for this study.

Conflict of interest

In the past three years, SML has received personal financial support from Janssen, Otsuka and Sunovion and research grant income from Janssen and Lundbeck in connection with therapeutic initiatives for psychosis. The authors VV, PK, KR, PS and ARS have no conflict of interest.

Bibliography

Adams, R.A., Stephan, K.E., Brown, H.R., Frith, C.D., Friston, K.J., 2013. The Computational Anatomy of Psychosis. Frontiers Psychiatry 4.

American Psychiatric Association, 2000. Diagnostic and statistical manual of mental disorders (4th ed., text rev. - DSM-IV-R). Washington, DC,

Andreasen, N.C., Pressler, M., Nopoulos, P., Miller, D., Ho, B.-C., 2010. Antipsychotic Dose Equivalents and Dose-Years: A Standardized Method for Comparing Exposure to Different Drugs.

Biological Psychiatry 67, 255–262.

Averbeck, B.B., Evans, S., Chouhan, V., Bristow, E., Shergill, S.S., 2011. Probabilistic learning and inference in schizophrenia. Schizophrenia Research 127, 115–122.

Chalk, M., Seitz, A.R., Seriès, P., 2010. Rapidly learned stimulus expectations alter perception of motion. Journal of Vision 10(8), 2.

Corlett, P.R., Frith, C.D., Fletcher, P.C., 2009a. From drugs to deprivation: a Bayesian framework for understanding models of psychosis. Psychopharmacology 206, 515–530.

Corlett, P.R., Simons, J.S., Pigott, J.S., Gardner, J.M., Murray, G.K., Krystal, J.H., Fletcher, P.C., 2009b. Illusions and delusions: relating experimentally-induced false memories to anomalous experiences and ideas. Frontiers Behavioural Neuroscience 3, 53.

Crawford, T.J., Hamm, J.P., Kean, M., Schmechtig, A., Kumari, V., Anilkumar, A.P., Ettinger, **U.**, 2010. The perception of real and illusory motion in schizophrenia. Neuropsychologia 48, 3121–3127.

Danion, J.-M., Meulemans, T., Kauffmann-Muller, F., Vermaat, H., 2001. Intact Implicit Learning in Schizophrenia. American Journal of Psychiatry 158, 944–948.

Daunizeau, J., Adam, V. and Rigoux, L., 2014. VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Computational Biology*, *10*(1), p.e1003441

Dima, D., Dietrich, D.E., Dillo, W., Emrich, H.M., 2010. Impaired top-down processes in schizophrenia: A DCM study of ERPs. NeuroImage 52, 824–832.

Dima, D., Roiser, J.P., Dietrich, D.E., Bonnemann, C., Lanfermann, H., Emrich, H.M., Dillo, W., 2009. Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. NeuroImage 46, 1180–1186.

Evans, S., Almahdi, B., Sultan, P., Sohanpal, I., Brandner, B., Collier, T., Shergill, S.S., Cregg, R., Averbeck, B.B., 2012. Performance on a probabilistic inference task in healthy subjects receiving ketamine compared with patients with schizophrenia. Journal of Psychopharmacology 26, 1211–1217.

Fiser, J., Berkes, P., Orbán, G., Lengyel, M., 2010. Statistically optimal perception and learning: from behavior to neural representations. Trends Cognitive Science (Regul Ed) 14, 119–130.

Fletcher, P.C., Frith, C.D., 2009. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nature Reviews Neuroscience 10, 48–58.

Freeman, D., Pugh, K., Garety, P., 2008. Jumping to conclusions and paranoid ideation in the general population. Schizophrenia Research 102, 254–260.

Friston, K.J., 2005. Hallucinations and perceptual inference. Behavioral and Brain Sciences 28, 764–766.

Friston, K.J., 2010. The free-energy principle: a unified brain theory? Nature Review Neuroscience 11, 127–138.

Friston, K.J., 2012. The history of the future of the Bayesian brain. NeuroImage 62, 1230–1233.

Friston, K., Brown, H.R., Siemerkus, J. and Stephan, K.E., 2016. The dysconnection hypothesis (2016). Schizophrenia research, 176(2), pp.83-94.

Frith, C.D., Friston, K.J., 2012. False perceptions and false beliefs: understanding schizophrenia. Working Group on Neurosciences and the Human Person: New Perspectives on Human Activities, The Pontifical academy of Sciences 8–10.

García-Pérez, M.A., 1998. Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. Vision Research. 38, 1861–1881.

Garety, P., Joyce, E.M., Jolley, S., Emsley, R., Waller, H., Kuipers, E., Bebbington, P., Fowler, D., Dunn, G., Freeman, D., 2013. Neuropsychological functioning and jumping to conclusions in delusions. Schizophrenia Research.

Garety PA, Freeman D (2013): The past and future of delusions research: from the inexplicable to the treatable. British Journal of Psychiatry 203: 327–333.

Gekas, N., Chalk, M., Seitz, A.R., Series, P., 2013. Complexity and specificity of experimentally-induced expectations in motion perception. Journal of Vision 13.

Gold, J.M., Hahn, B., Strauss, G.P., Waltz, J.A., 2009. Turning it upside down: areas of preserved cognitive function in schizophrenia. Neuropsychological Reviews 19, 294–311.

Horton, H.K., Silverstein, S.M., 2011. Visual Context Processing Deficits in Schizophrenia: Effects of Deafness and Disorganization. Schizophrenia Bulletin 37, 716–726.

Huq, S.F., Garety, P.A., Hemsley, D.R., 1988. Probabilistic judgments in deluded and non-deluded subjects. Quarterly Journal of Experimental Psychology A 40, 801–812.

Jardri, R., Denève, S., 2013. Circular inferences in schizophrenia. Brain.

Jardri, R., Duverne, S., Litvinova, A.S. and Denève, S., 2017. Experimental evidence for circular inference in schizophrenia. Nature communications, 8, p.14218.

Kaas, R.E., Raftery A.E., 1995. Bayes Factors. Journal of the American Statistical Association 90(430), 773-795.

Karvelis, P., Seitz, A.R., Lawrie, S.M. and Seriès, P., 2018. Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration. eLife, 7, p.e34115.

Keane, B.P., Silverstein, S.M., Wang, Y., Papathomas, T.V., 2013. Reduced depth inversion illusions in schizophrenia are state-specific and occur for multiple object types and viewing conditions. Journal of Abnormal Psychology 122, 506–512.

Kéri, S., Kelemen, O., Szekeres, G., Bagóczky, N., Erdélyi, R., Antal, A., Benedek, G., Janka, Z., 2000.

Schizophrenics know more than they can tell: probabilistic classification learning in schizophrenia. Psychological Medicine 30, 149–155.

Knill, D.C., Pouget, A., 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neuroscience 27, 712–719.

Kok, P., Brouwer, G.J., van Gerven, M.A.J., de Lange, F.P., 2013. Prior Expectations Bias Sensory Representations in Visual Cortex. Journal of Neuroscience 33, 16275–16284.

Laquitaine S., Gardner J., 2018. A Switching Observer for Human Perceptual Estimation. Neuron 97(2), pp.462-474

Leucht, S., Samara, M., Heres, S., Patel, M.X., Furukawa, T., Cipriani, A., Geddes, J. and Davis, J.M., 2015. Dose equivalents for second-generation antipsychotic drugs: the classical mean dose method. Schizophrenia bulletin, 41(6), pp.1397-1402.

Levitt, H., 1971. Transformed up-down methods in psychoacoustics. Journal of Acoustical Society of America. 49, Suppl 2:467–.

Mardia, K.V., 1972. Statistics of directional data, Academic Press, Inc. (London) LTD.

Marvel, C.L., Schwartz, B.L., Howard, D.V., Howard, J.H., 2005. Implicit learning of non-spatial sequences in schizophrenia. Journal of the International Neuropsychological Society: JINS 11, 659–667.

Nelson, H. E., & Willison, J. R. (1991). The National Adult Reading Test (NART): test manual (2nd ed.). Windsor: NFER-Nelson

Notredame CE, Pins D, Denève S 2014: What Visual illusions teach us about schizophrenia. Frontiers in Integrative physiology

Nuechterlein, K.H., 1977. Reaction time and attention in schizophrenia. Schizophrenia Bulletin 3.

Palminteri, S., Wyart, V., and Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. Trends in Cognitive Sciences, 21(6): 425-433.

Powers, A.R., Mathys, C. and Corlett, P.R., 2017. Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. Science, 357(6351), pp.596-600.

Rigoux, L., Stephan, K.E., Friston, K.J. and Daunizeau, J., 2014. Bayesian model selection for group studies—revisited. *Neuroimage*, *84*, pp.971-985.

Sato, Y. and Kording, K.P., 2014. How much to trust the senses: Likelihood learning. Journal of Vision, 14(13), pp.13-13.

Seriès, P., Seitz, A.R., 2013. Learning what to expect (in visual perception). Frontiers in human neuroscience 7.

Schmack, K., de Castro, A.G.C., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J.D., Heinz, A., Petrovic, P. and Sterzer, P., 2013. Delusions and the role of beliefs in perceptual inference. Journal

of Neuroscience, 33(34), pp.13701-13712.

Schmack, K., Schnack, A., Priller, J. and Sterzer, P., 2015. Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. Schizophrenia Research: Cognition, 2(2), pp.72-77.

Schmack, K., Rothkirch, M., Priller, J. and Sterzer, P., 2017. Enhanced predictive signalling in schizophrenia. Human brain mapping, 38(4), pp.1767-1779.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. CRC Press.

Silverstein SM, Keane BP (2011a): Vision science and schizophrenia research: toward a re-view of the disorder. Editors' introduction to special section. Schizophrenia Bulletin 37: 681–689.

Silverstein SM, Keane BP (2011b): Perceptual organization impairment in schizophrenia and associated brain mechanisms: review of research from 2005 to 2010. Schizophrenia Bulletin 37: 690–699.

Skottun, B.C., Skoyles, J.R., 2007. Contrast sensitivity and magnocellular functioning in schizophrenia. Vision Research. 47, 2923–2933.

Speechley, W.J., Whitman, J.C., Woodward, T.S., 2010. The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. Journal of Psychiatry Neuroscience 35, 7–17.

Sterzer, P., Adams R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli L., Petrovic, P., Uhlhaas, P., Voss, M., Corlett, P.R., 2018. The predictive coding account of psychosis. Biological Psychiatry 84:634-643.

Tschacher, W., Schuler, D., Junghan, U., 2006. Reduced perception of the motion-induced blindness illusion in schizophrenia. Schizophrenia Research 81, 261–267.

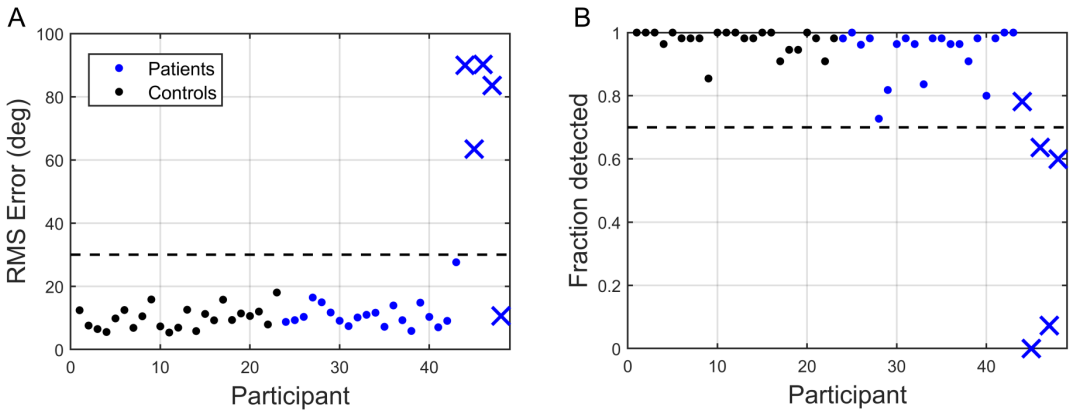
Valton, V., Romaniuk, L., Steele, J.D., Lawrie, S. and Seriès, P., 2017. Comprehensive review: Computational modelling of schizophrenia. Neuroscience & Biobehavioral Reviews. 83:631-646

Wand, M.P., Jones, M.C., 1994. Kernel Smoothing. CRC Press.

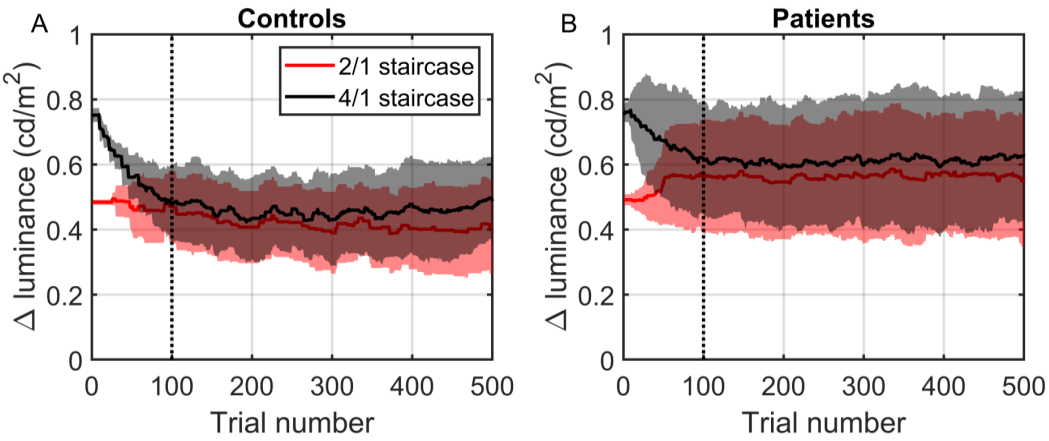
Wechsler, D. (1999). Wechsler Abbreviated Scale of Intelligence. The Psychological Corporation: Harcourt Brace & Company. New York, NY.

Acquisition of visual priors and induced hallucinations in chronic schizophrenia: Supplementary information

V. Valton, P. Karvelis, K. L. Richards, A. R. Seitz, S. M. Lawrie, P. Seriès



Supplementary Figure 1: Performance on high contrast trials. **(A)** Root mean square error of estimations. **(B)** Fraction of trials on which the stimulus was detected. The dots represent included participants; the cross marks represent excluded participants. Dashed lines denote inclusion criteria (30 degree estimation and 70% detection).

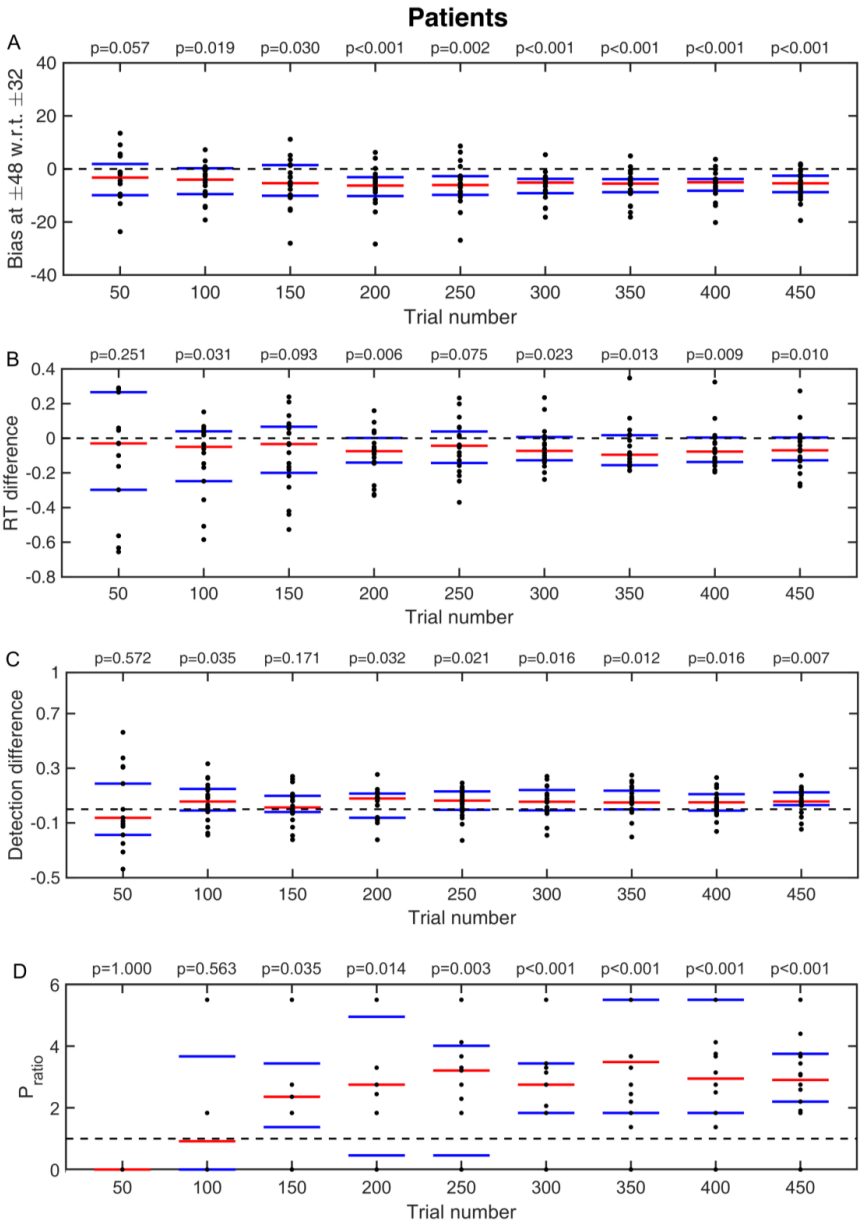


Supplementary Figure 2: Convergence of 2/1 and 4/1 staircase luminance levels. **(A)** Controls, **(B)** Patients. Both groups reached convergence after ~100 trials. These trials were excluded from further data analysis.

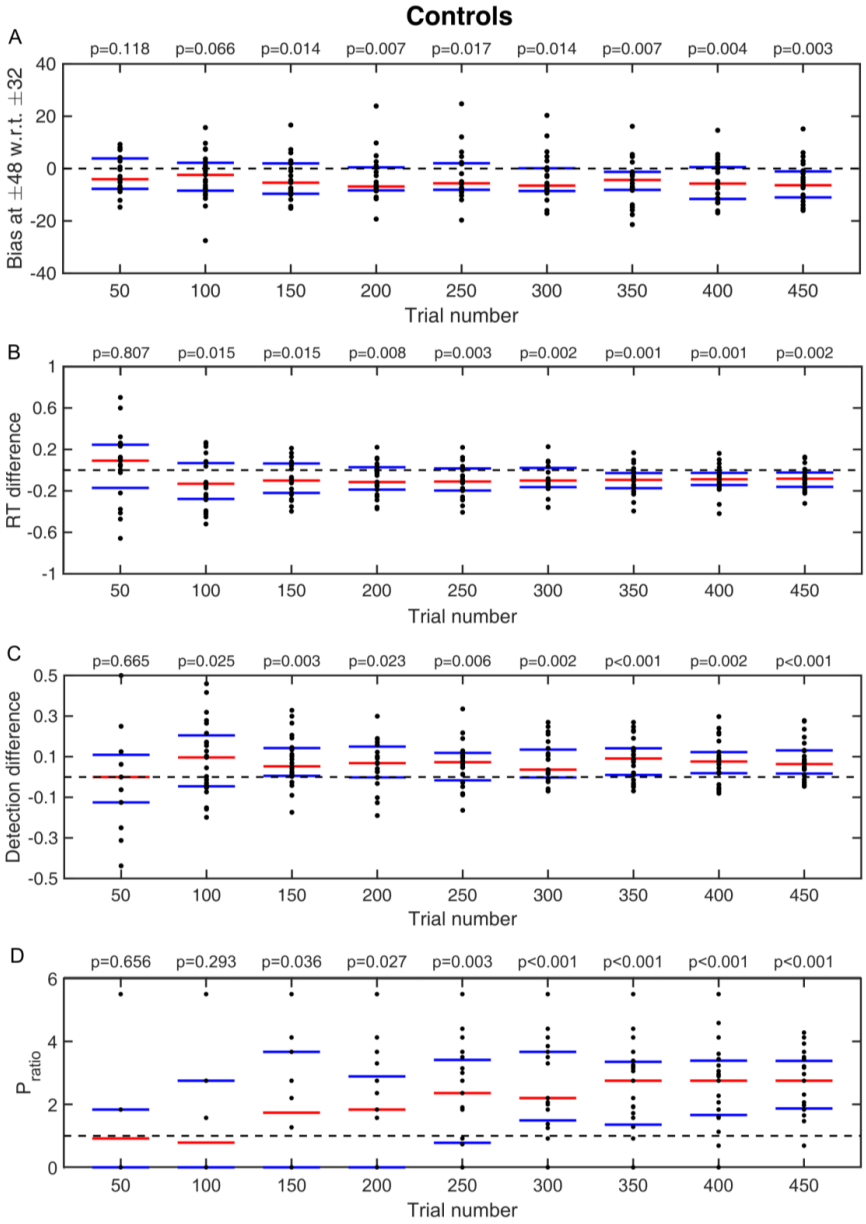
Emergence of prior effects

We wanted to determine when the acquired expectations started to have a significant effect on performance and whether this was the same for both groups. First and most importantly, we wanted to know when the estimations on low contrast trials became biased towards ±32°. To do so, we computed cumulative moving averages at every 50 trials for the bias at ±48° with respect to bias at ±32°. Next, we looked at estimation reaction times (RT) on low contrast trials and compared mean RT of each individual at ±32° with mean RT at all other directions. Similarly, we looked at the average detection performance on low contrast trials and compared the fraction of trials in which stimulus was detected at ±32° with the mean fraction detected over all other presented directions.

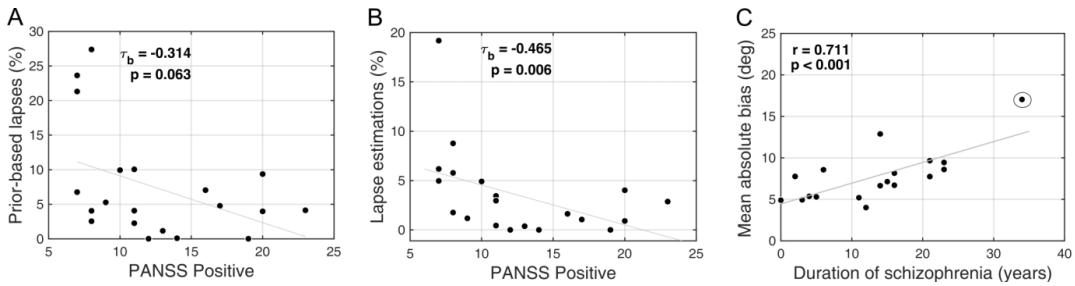
We found that both patients and controls showed signs of very rapid acquisition of the priors. For the bias to become statistically significant, it took less than 100 trials for patients (**Supplementary Fig. 3A**) and less than 150 trials for controls (**Supplementary Fig. 4A**). Similarly, both patients and controls became significantly faster and significantly better at detecting stimulus moving at ±32° within the first 100 trials (**Supplementary Fig. 3B, C and 4B, C respectively**).



Supplementary Figure 3: Emergence of prior effects in patient group. **(A)** Cumulative moving averages of bias at ±48° with respect to bias at ±32°. **(B)** Cumulative moving averages of median differences between estimation RTs at ±32° and RTs at all other directions. **(C)** Cumulative moving averages of median differences between fraction of detected stimuli at ±32° and fraction detected at all other directions. **(D)** Cumulative moving averages of the probability ratio of hallucinating predominantly around ±32 on no-stimulus trials. Red bars indicate median values and blue bars indicate 25th and 75th percentiles. p-values are denoted above each plot (one-tailed Wilcoxon signed rank test).



Supplementary Figure 4: Emergence of prior effects in control group. **(A)** Cumulative moving averages of bias at ±48° with respect to bias at ±32°. **(B)** Cumulative moving averages of median differences between estimation RTs at ±32° and RTs at all other directions. **(C)** Cumulative moving averages of median differences between fraction of detected stimuli at ±32° and fraction detected at all other directions. **(D)** Cumulative moving averages of the probability ratio of hallucinating predominantly around ±32 on no-stimulus trials. Red bars indicate median values and blue bars indicate 25th and 75th percentiles. p-values are denoted above each plot (one-tailed Wilcoxon signed rank test).



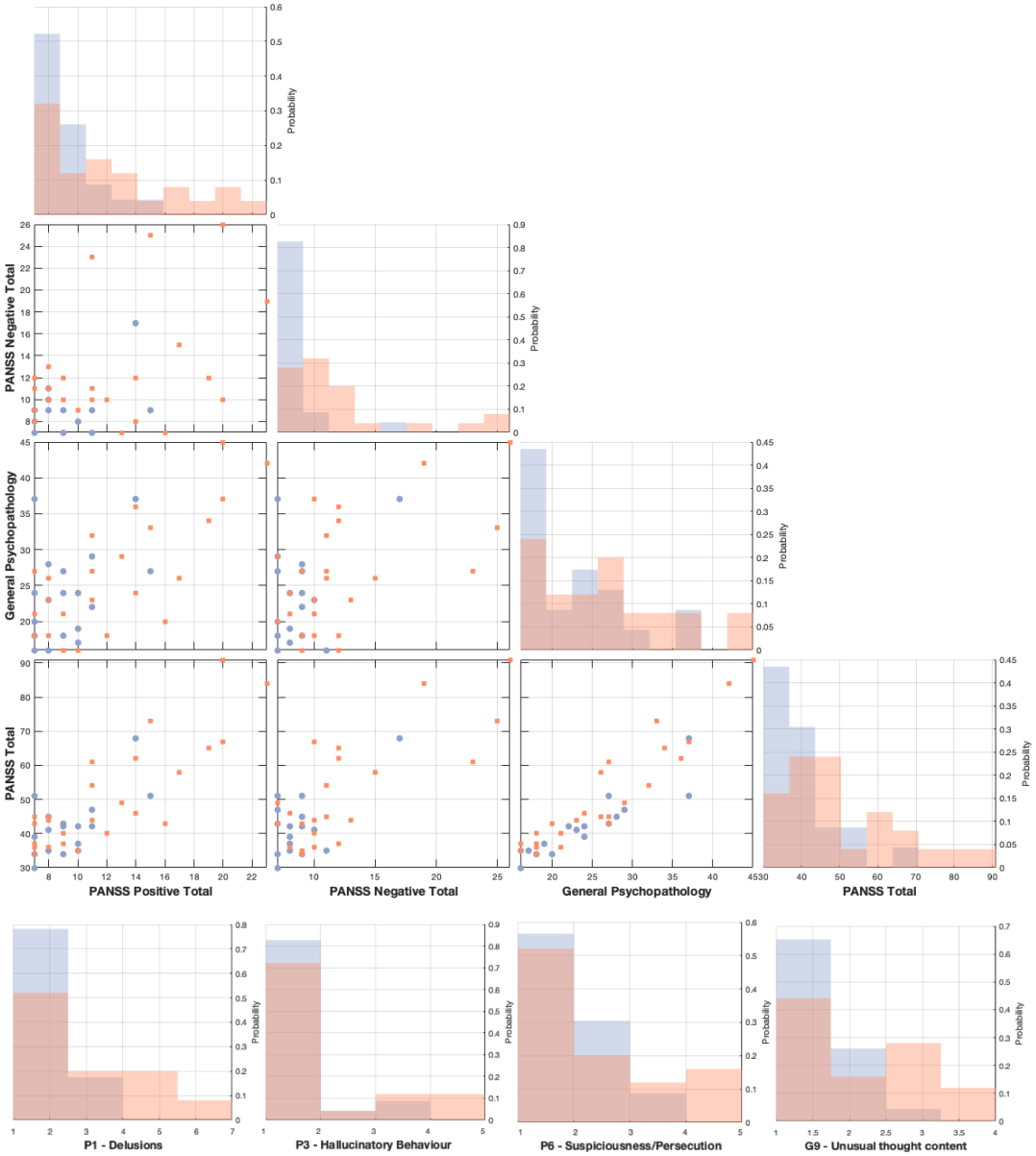
Supplementary Figure 5: Exploratory analysis. (**A**) Prior-based lapses (estimated via BAYES_P) as a function of PANSS positive score (τ_b = -0.314, p = 0.063; Kendall's correlation). (**B**) Lapse estimations (determined via Eq.(1)) and PANSS positive score (τ_b = -0.465, p = 0.006; Kendall's correlation). (**C**) Mean absolute bias as a function of duration of illness. The plot shows a positive relationship between the duration of illness and the amount of bias exhibited during the task (r = 0.711, p < 0.001; Pearson's correlation. The correlation remained highly significant after excluding the outlier (circled) (r = 0.523, p = 0.003; Pearson's correlation).

Detailed clinical characteristics

The demographic and clinical characteristics of the included participants can be seen in Table S1 and Supplementary Figure 6. It is of note that patients had relatively low levels of symptoms on the PANSS, but ongoing functional impairment. Most were assigned a PANSS rating of 1 (absent), 2 (minimal/questionable) or 3 (mild) for delusions, hallucinatory behaviour, suspiciousness/persecution, and unusual thought content, reflecting the clinically stable state of the psychosis group. Except for one patient, all had been taking anti-psychotic medication at the time of testing (atypical, $n = 18$; typical, $n = 1$). Nine were also taking anti-depressants and two were taking sodium valproate.

	Patients (n=20)	Controls (n=23)	*Statistic*	p-value
Delusions (P1)	2.60 (1.90)	1.48 (0.79)	$W = 158$	< .05
Hallucinatory behaviour (P3)	1.85 (1.31)	1.22 (0.60)	$W = 175$	.07
Suspiciousness/Persecution (P6)	2.00 (1.26)	1.61 (0.84)	$W = 198$	.40
Unusual thought content (G9)	2.05 (1.10)	1.35 (0.57)	$W = 150$	< .05

Table S1. Demographic and detailed Clinical Characteristics (mean and standard deviation in parentheses). P# refer to specific items/symptoms of the PANSS (Positive and Negative Symptom Scale). For all measures Wilcoxon rank-sum tests (normal approximation with continuity correction) were used.



Supplementary Figure 6: Histogram and pair-plots for general PANSS subscales, and histograms for items P1, P3, P6 and G9 of the PANSS. Blue data represents controls, while red data represents patients. The graphical overlay displaying the symptom data for patients and controls suggests that the patient groups is homogeneous and relatively well with respect to symptom severity when compared to the control group.

Normality and Non-parametric testing

We ran normality tests among repeated measures for each group separately, as well as for each group averaged across angles (**Table S2**).

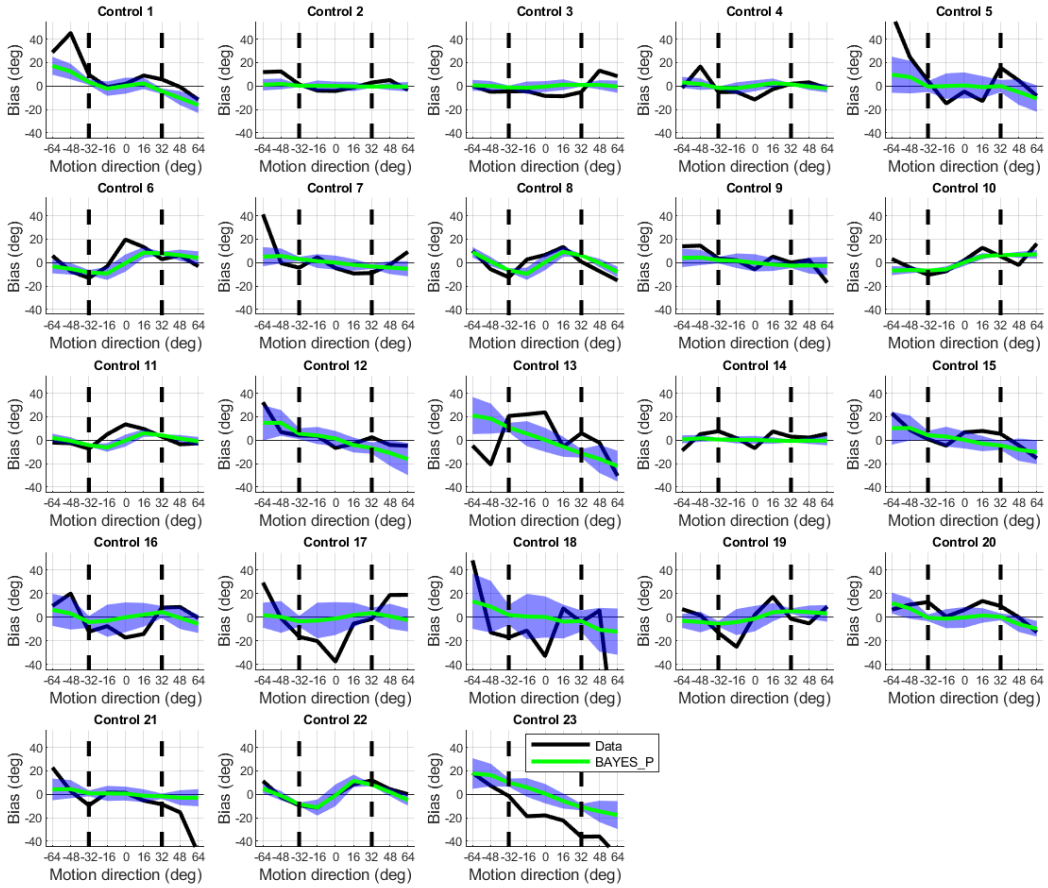
Angle	Group	Bias		Variability		Lapses		RT		Detection	
		SW Stat.	p-value	SW Stat.	p-value	SW Stat.	p-value	SW Stat.	p-value	SW Stat.	p-value
0°	*Patients*	.964	.616	.944	.290	.236	**<.001**	.963	.602	.964	.622
	Controls	.953	.337	.929	.103	.486	**<.001**	.968	.632	.960	.472
±16°	*Patients*	.914	.075	.931	.163	.530	**<.001**	.963	.602	.970	.757
	Controls	.967	.614	.953	.341	.705	**<.001**	.984	.966	.861	**.004**
±32°	*Patients*	.899	**.039**	.883	**.020**	.906	.053	.831	**.003**	.927	.133
	Controls	.904	**.030**	.915	.053	.860	**.004**	.951	.300	.939	.170
±48°	*Patients*	.839	**.003**	.951	.385	.505	**<.001**	.959	.527	.982	.956
	Controls	.922	.074	.866	**.005**	.847	**.002**	.953	.340	.963	.531
±64°	*Patients*	.702	**<.001**	.897	**.036**	.740	**<.001**	.958	.503	.934	.183
	Controls	.790	**<.001**	.901	**.027**	.861	**.004**	.963	.519	.972	.731
Mean	*Patients*	.583	**<.001**	.948	.340	.682	**<.001**	.894	**.033**	.919	.096
	Controls	.743	**<.001**	.937	.153	.886	**.015**	.967	.626	.933	.116

Table S2. Shapiro-Wilk normality tests for estimation bias, variability, lapses, reaction time and detection rate at each stimulus motion direction (0°, ±16°, ±32° ±48° ±64°) and averaged across motion directions (mean) for each group (patients and controls). Degrees of freedom are df = 20 for patients and df = 23 for controls for all angles. Normality violations below $\alpha = 0.05$ significance level are highlighted in bold.

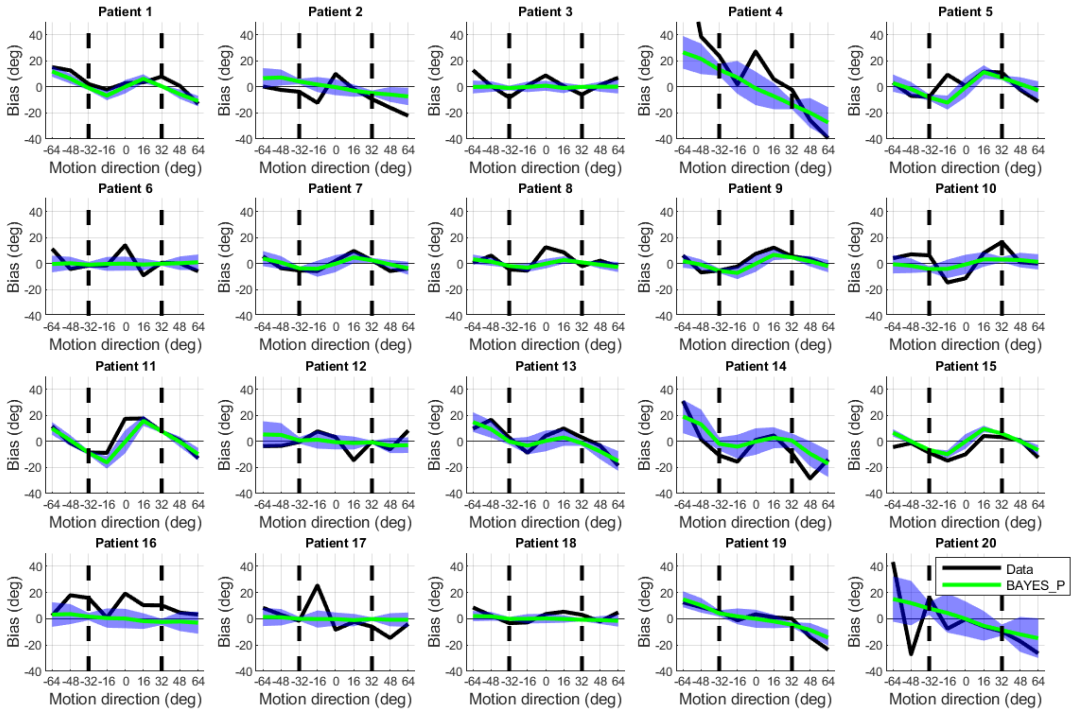
Some deviations from normality were identified for estimation biases and lapses. For these measures we re-ran group comparisons using two-sided Wilcoxon rank-sum test that does not require normality assumptions. To our knowledge there is no well-established non-parametric equivalent to the 2-way mixed ANOVA, therefore, we focused on verifying group comparisons on the mean values of bias and lapses. Most importantly, while being more conservative than parametric tests, the non-parametric rank-sum test results were found to be in complete agreement with the ANOVA test results: patients exhibited significantly less lapse estimations than controls ($W = 324$, $p = 0.005$, two-sided rank-sum test) and there were no differences in the mean absolute bias between the groups ($W = 399$, $p = 0.324$, two-sided rank-sum test). Furthermore, we also report Bayesian statistical analyses on all of the measures reported, which also confirm the frequentist parametric and non-parametric analyses findings. Therefore, we

feel confident that the findings reported in the manuscript do not suffer from the normality assumptions being violated.

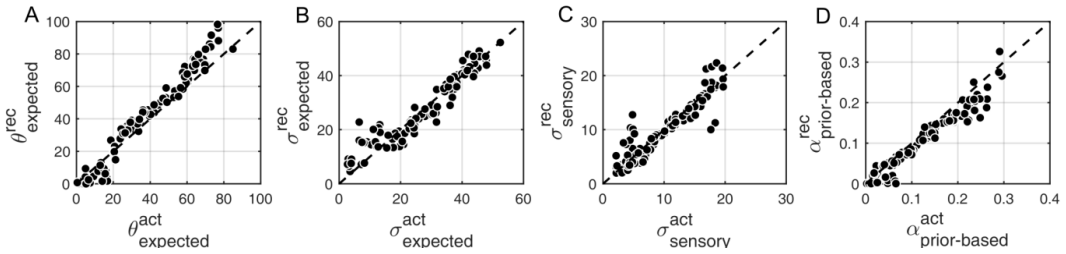
It is worth noting however that the effects of deviations from normality assumptions in ANOVAs has been well studied in the literature, and such tests were found to be robust to normality assumptions violations which could explain our findings being replicated in both Bayesian and non-parametric tests (Glass et al., 1972; Harwell et al., 1992; Lix et al., 1996).



Supplementary Figure 7: BAYES_P bias fits for each subject in the control group. Black solid line is the behavioural data, green line is the mean bias and blue shaded area is the standard deviation of the bias as predicted by BAYES_P. The model predictions were generated using the parameter values obtained from maximum likelihood estimation. The model was simulated to produce the same number of responses as in the empirical data of each subject. For each subject, the simulation was repeated 30 times to obtain a distribution of biases, which then was used to calculate the mean and the standard deviation of the bias.



Supplementary Figure 8: BAYES_P bias fits for each subject in the patient group. Black solid line is the behavioural data, green line is the mean bias and blue shaded area is the standard deviation of the bias as predicted by BAYES_P. The model predictions were generated using the parameter values obtained from maximum likelihood estimation. The model was simulated to produce the same number of responses as in the empirical data of each subject. For each subject, the simulation was repeated 30 times to obtain a distribution of biases, which then was used to calculate the mean and the standard deviation of the bias.



Supplementary Figure 9. Parameter recovery with BAYES_P model. **(A)** $\theta_{expected}$ - mean of the prior expectations ($R^2 = 0.96$), **(B)** $\sigma_{expected}$ - uncertainty of the prior distribution ($R^2 = 0.89$), **(C)** $\sigma_{sensory}$ - uncertainty in the sensory likelihood ($R^2 = 0.84$), **(D)** $\alpha_{prior\text{-}based}$ – prior-based lapse rate ($R^2 = 0.94$). X-axes – actual parameters used for simulating the data (denoted with the superscript 'act'), Y-axes – recovered parameters (denoted with the superscript 'rec') from fitting the model to the simulated data. The dashed diagonal line is a reference line indicating perfect parameter recovery.

Modelling

Variant Bayes models

The model 'Bayes' is similar to the model 'Bayes_P' described in the main text, but has uniform lapse estimations instead of 'prior-based' lapse estimations, such that:

$$p(\theta_{estimate}|\theta_{perceived}) = (1 - \alpha) \cdot V(\theta_{perceived}, \sigma_{motor}) + \alpha/2\boldsymbol{\pi}$$

The model 'Bayes_var' is similar to the model 'Bayes', except that to account for the possibility of exposure effect to sensory uncertainty, we allow sensory uncertainty $\sigma_{sensory}$ to vary for each of the following angles 0°, ±16°, ±32°, ±48°, ±64°, resulting in five $\sigma_{sensory}$ free parameters.

The model 'Bayes_varmin' is similar to the model 'Bayes', except that we allow $\sigma_{sensory}$ to vary only for the most presented angles ±32°, resulting in two $\sigma_{sensory}$ free parameters ($\sigma_{sensory}$ at ±32°, and $\sigma_{sensory}$ at all other angles).

Response strategy models

We wanted to control for the possibility that the task behaviour might be explained by simple behavioural strategies that do not involve Bayesian integration. This class of models assumed that participants did not combine their expectations with sensory information, but relied on either of them alone on any given trial.

The first model, 'ADD1', assumed that estimations derived from prior expectations were simply sampled from a learnt prior distribution, $p_{expected}(\theta)$, which was parameterized as in Eq (4). However, on trials when participants did perceive motion direction, it was based solely on the sensory input, $p_{sensory}(\theta_{sensory}|\theta_{actual}) = V(\theta_{actual}, \sigma_{sensory})$.

Putting together the estimations derived from sensory input and the ones derived from learnt expectations, and the possibility of random estimations, the average distribution of estimation responses for a single participant is:

$$p(\theta_{estimate}|\theta_{actual}) = (1-\alpha)\cdot [(1-a(\theta)) \cdot p_{sensory}(\theta_{sensory}|\theta_{actual}) + a(\theta) \cdot p_{expected}(\theta_{estimate})] * V(0,\sigma_{motor}) + \alpha, \quad (9)$$

where $a(\theta)$ determines the proportion of trials in which participants sampled from the acquired prior, $p_{expected}(\theta)$; asterisk ($*$) denotes convolution. The resulting 'ADD1' model has 9 free parameters ($\theta_{expected}$, $\sigma_{expected}$, $a(\theta)$ (which can take a different value for each of the 5 angles: 0°, $\pm 16^\circ$, $\pm 32^\circ$, $\pm 48^\circ$, $\pm 64^\circ$), $\sigma_{sensory}$ and α).

The second model, 'ADD2', was the same as 'ADD1' except that it had more complex strategy for trials when participants relied on the prior: instead of sampling from the complete acquired prior distribution ranging from -180° to $+180^\circ$ (Eq. (4)), they sampled from only one half of it, negative ($-180°$ to $0°$) or positive ($0°$ to $+180°$), depending on which side of the distribution the actual stimulus occurred on:

$$p_{expectedN}(\theta) = V(-\theta_{expected}, \sigma_{expected}) \quad (10)$$

$$p_{expectedP}(\theta) = V(\theta_{expected}, \sigma_{expected}) \quad (11)$$

Incorporating this into the distribution of estimation responses results in:

$$\begin{aligned} p(\theta_{estimate}|\theta_{actual}) = (1-\alpha)\cdot[((1 - a(\theta) - b(\theta)) \cdot p_{sensory}(\theta_{senssory}|\theta_{actual}) \\ + a(\theta) \cdot p_{expectedN}(\theta_{estimate}) \\ + b(\theta) \cdot p_{expectedP}(\theta_{estimate})] \\ * V(0,\sigma_{motor}) + \alpha, \end{aligned} \quad (12)$$

where asterisk (∗) denotes convolution; a(θ) and b(θ) determine the proportion of trials in which participants sample from either negative or positive parts of the prior distribution, respectively.

Finally, we also considered two variations of the 'ADD1' and 'ADD2' models. These were identical to 'ADD1' and 'ADD2' except from setting $\sigma_{expected}$ to zero (i.e. no uncertainty); that is, on trials when perceptual estimates were derived only from expectations, they were equal to the mode of the learnt distribution. These models are referred to as 'ADD1_m' and 'ADD2_m'.

Parameter estimation

We used the performance in trials with the highest contrast level to estimate motor noise, σ_{motor}, for each individual. We assumed that at this level sensory uncertainty was close to zero ($\sigma_{sensory} \approx 0$). The motor noise was determined by fitting estimation responses at the highest contrast level to the distribution in Eq. (2) using the actual motion direction, θ_{actual}, as the mean. The estimated motor noise was used in all the subsequent model fitting as a fixed parameter.

The rest of the free parameters of each model were estimated by fitting the response data from the two staircased contrast levels (~200 trials per participant). For each model with a set of free parameters M, we computed the probability distribution $p(\theta_{estimate}|\theta_{actual}; M)$ of making an estimate $\theta_{estimate}$ given the actual stimulus direction θ_{actual}. For the response strategy models, by

definition, the $p(\theta_{estimate}|\theta_{actual}; M)$ corresponds to average behaviour in the task (Equations 9 and 12). Bayesian models, on the other hand, explicitly model trial-to-trial variability in the posterior estimate, which in our case is the mean of the posterior (Eq. (6)). To relate this to the behavioural data we built a distribution of 1,000 samples for each presented angle (where each sample is the mean of the posterior obtained via Eq. (6) and perturbed by motor noise via Eq. (7) or (8)).

The parameters were estimated by maximizing the fit of the log likelihood function for the experimental data for each participant individually:

$$M = \underset{M}{\operatorname{argmax}}\left[\sum_i^n \log\left(\mathrm{p}\left(\theta_{estimate} = \theta_{i,data}|\theta_i\right)\right)\right], \quad (13)$$

where $\theta_{i,data}$ is participant's estimation response, θ_i is the actual presented motion direction on the ith trial and n is the number of trials. The maximum likelihood was found using *fminsearchbnd* function in Matlab, by minimizing negative log-likelihood. Parameters α, $a(\theta)$ and $b(\theta)$ were bounded between 0 and 1, while $\theta_{expected}$, $\sigma_{expected}$ and $\sigma_{sensory}$ were bounded from 0 to ∞. To reduce the possibility of convergence at a local maxima we constructed a grid of initial $\sigma_{expected}$ and $\sigma_{sensory}$ parameter values covering the range as found in previous studies ($\sigma_{expected}$ from 10° to 25° in 3° increments and $\sigma_{sensory}$ from 5° to 15° in 3° increments, which resulted in 22 different initializations). A set of parameters with the largest log-likelihood was selected as the best fit.

Model Comparison

To compare the model fits we used Bayesian Information Criterion (BIC), which approximates the log of model evidence (e.g., see Burnham and Anderson, 2004):

$$-2 \cdot \log(P(D|M)) \approx BIC = -2 \cdot \log(P(D|M, \Theta)) + k \cdot \log(n) \quad , \quad (14)$$

where M is model, D is observed data and P (D|M, Θ) is the likelihood of generating the experimental data given the most likely set of parameters, Θ; k is the number of model parameters and n is the number of data points (or equivalently, the number of trials). BIC evaluates the model by how it fits the data by also penalizing for the number of parameters (i.e. model complexity) to avoid over-fitting. Lower BIC score indicates a better model. We also ran a random effect Bayesian model selection analysis for group studies (Rigoux et al., 2014). We used the VBA Matlab toolbox (Daunizeau et al., 2014) to perform this analysis, and used participant-level BIC as an approximation of the log-model evidence

Parameter recovery

To test the reliability of the parameter estimates of our winning model we performed parameter recovery. This allowed us to simultaneously test whether parameters are identifiable (e.g., whether likelihood and prior uncertainty is not correlated and can be distinguished) and whether having ~200 trials (the amount of low contrast trials in our data) for data fitting and using maximum likelihood estimation are sufficient to give reliable results.

First, we generated 100 sets of parameters (i.e. 100 synthetic individuals) by randomly sampling each parameter from a Gaussian distribution which had a mean and variance as the parameter estimates from the collected participant data. Second, for each set of parameters we simulated data for 200 trials with the winning model by randomly sampling from the estimation probability distribution, which, as for the behavioural data, was built from a 1000 posterior means (Eq. (6)), each perturbed by motor noise (Eq. (8)) Finally, we fitted the winning model to the simulated data. To evaluate the goodness of recovered parameters we computed the coefficient of determination (R^2) for a linear regression, which quantified how well the actual parameters predicted the recovered ones.

Power calculations

Power calculations (a-priori): detecting effect of the prior on behaviour

We performed power calculations using an independent dataset from a previously published study that used the exact same task design but had 2 sessions of 850 trials (Chalk *et al*., 2010). We used the first 567 trials of the first session for the power analysis, to conform to the same trial structure present in the current study.

'Bias' and 'Detection rate' were selected as the behavioural measures relevant for determining the effect of the prior on behaviour. Given these parameters, power calculations for 'Bias' and 'Detection rate' revealed that 18-20 subjects are required respectively to detect an effect with 80% power.

Power calculations (a-priori): detecting group differences

To our knowledge, the current task design has never been tested in a chronic Schizophrenia sample. We therefore turned to the literature to estimate the effect size between patients and controls in motion perception (Chen, 2011) and illusions (Dima et al., 2009; Dima et al., 2011).

We determined that 20 participants in each group should allow us to detect an effect size of at least 0.9 with 80% power, which is a smaller effect than previous effect sizes reported in motion perception (Chen et al., 1999a; Chen et al., 1999b), perceptual illusions (Dima et al., 2009; Dima et al., 2011), or cognition generally (Fioravanti *et al.*, 2012).

Based on this analysis we aimed to recruit approximately 25 subjects per group, which would permit to reliably detect the effect of: the prior on perception, as well as differences in motion perception or illusions between groups, even accounting for a potentially large (e.g. 20%) exclusion rate based on task performance.

References

Burnham, K.P. and Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociological methods & research, 33(2), 261-304.

Chalk, M., Seitz, A.R., Seriès, P., 2010. Rapidly learned stimulus expectations alter perception of motion. Journal of Vision 10, 2.

Chen, Y., Nakayama, K., Levy, D.L., Matthysse S., Holzman P.S. 1999. Psychophysical isolation of a motion-processing deficit in schizophrenics and their relatives and its association with impaired smooth pursuit. Proc Natl Acad Sci U S A. 96(8): 4724–4729.

Chen, Y., Palafox, G. P., Nakayama, K., Levy, D. L., Matthysse, S., Holzman P. S. 1999. Motion perception in schizophrenia. Arch Gen Psychiatry. 1999 Feb; 56(2):149-54.

Daunizeau, J., Adam, V. and Rigoux, L., 2014. VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Computational Biology*, *10*(1), p.e1003441Dima, D., Dietrich, D.E., Dillo, W., Emrich, H.M.. 2010. Impaired top-down processes in schizophrenia: A DCM study of ERPs. NeuroImage 52, 824–832.

Dima, D., Roiser, J.P., Dietrich, D.E., Bonnemann, C., Lanfermann, H., Emrich, H.M., Dillo, W., 2009. Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. NeuroImage 46, 1180–1186.

Fioravanti M, Bianchi V, Cinti ME. Cognitive deficits in schizophrenia : an updated meta-analysis of the scientific evidence. BMC Psychiatry. 2012; 12:64

Glass, G.V., P.D. Peckham, and J.R. Sanders. 1972. Consequences of failure to meet assumptions underlying fixed effects analyses of variance and covariance. Rev. Educ. Res. 42: 237-288.

Harwell, M.R., E.N. Rubinstein, W.S. Hayes, and C.C. Olds. 1992. Summarizing Monte Carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases. J. Educ. Stat. 17: 315-339.

Lix, L.M., J.C. Keselman, and H.J. Keselman. 1996. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. Rev. Educ. Res. 66: 579-619.

Rigoux, L., Stephan, K.E., Friston, K.J. and Daunizeau, J., 2014. Bayesian model selection for group studies—revisited. *Neuroimage*, *84*, pp.971-985.

4

Perceptual Bayesian inference in people with ASD diagnosis

This chapter includes a manuscript to be submitted for publication: Karvelis, P., Richards, K. L., Seitz, A. R., Lawrie, S. M., Seriès, P. (in prep.) Intact Bayesian inference in visual motion perception in adults with ASD.

Intact Bayesian inference in visual motion perception in adults with ASD

Povilas Karvelis[1], Katie L. Richards[2], Aaron R. Seitz[3], Stephen M. Lawrie[2,4], Peggy Seriès[1,*]

[1] Institute for Adaptive and Neural Computation, University of Edinburgh, UK

[2] Department of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, UK

[3] Department of Psychology, University of California Riverside, CA, USA

[4]Patrick Wild Centre, University of Edinburgh, Edinburgh, UK

**Correspondence should be addressed to: pseries@inf.ed.ac.uk*

Abstract

In Bayesian inference framework, perception is an active process where incoming ambiguous sensory information ('likelihood') is integrated with accumulated environmental statistics ('prior'). Current computational theories of Autism Spectrum Disorders (ASD) suggest that many of the sensory symptoms could be understood within such framework: perception being less influenced by the priors. We investigated whether adults with ASD (N = 17) differed from age-, gender- and IQ-matched controls (N = 30) in rapid acquisition of perceptual priors and their influence on perception. We used a visual motion estimation task designed for quantitatively assessing both prior and likelihood distributions. While ASD group was faster in their responses, they showed no differences in prior acquisition nor in precision of their sensory likelihoods. This contrasts with our previous findings using the same task in the general population, where we found autistic traits to be associated with more precise likelihoods.

Introduction

Perception can be thought of as Bayesian inference, where ambiguous sensory information ('likelihood') is integrated with prior knowledge of the world ('prior') to infer the causes of sensory signals (Rao & Ballard, 1999; Knill & Pouget, 2004; Friston, 2005). Such priors are continuously updated to match long-term and short-term statistical regularities of environmental features (see Seriès & Seitz, 2010 for a review). Importantly, the extent to which priors shape perception depends on the relative amount of precision with which priors and likelihoods are represented in the brain.

Recent theoretical accounts hypothesize that autism spectrum disorders (ASD) could be understood in terms of imbalanced weighting of priors and likelihoods, resulting in perception that is closer to the sensory input. This could be due to priors being generally weaker - i.e. less precise (Pellicano & Burr, 2012), or due to sensory likelihoods being overly precise (Brock, 2012). Considering the hierarchical nature of perceptual inference and the dynamics of prior updating, overly precise likelihood might arise from insufficient sensory attenuation via top-down gain control (Lawson et al., 2014), and would lead to an inflexibly high rate of prior updating (Van de Cruys et al., 2014). These proposals can retrospectively explain a wide range of symptoms and experiment-based characterizations of ASD behaviour (for a review see Palmer et al., 2017).

However, experimental studies aimed at explicitly testing these Bayesian accounts so far have produced mixed results. A handful of studies found support for the weaker prior hypothesis: along ASD traits in the general population, Powell et al. (2016) reported weaker slow-speed priors in motion perception, while Skewes et al. (2015) found weaker priors in a probabilistic reasoning game involving orientation discrimination. In children with ASD, Karaminis et al. (2016) found a reduced central tendency effect in temporal interval reproduction. However, plenty of other studies show priors in ASD to be intact: convexity and shape familiarity priors in figure-ground segregation in children and young adults with ASD (Spanò et al., 2016), the direction of gaze priors in high functioning adults with ASD (Pell et al., 2016), the light-from-above priors in children with ASD (Croydon et al., 2017), induced priors when disambiguating

two-tone (Mooney) images along ASD traits in the general population (Van de Cruys et al., 2018; although see Król & Król, 2019) and priors underlying the illusions of Kanizsa's Triangle in adolescents with ASD (Utzerath et al., 2019). A few studies have also investigated the dynamics of prior updating. Lawson et al. (2017) used a probabilistic associative learning task and found that adults with ASD estimated volatility to be more variable, which lead to smaller adjustments of their learning rate about stimuli contingencies when transitioning from stable to volatile phases of the task; however, the average learning rates were not found to be different. In contrast, a recent study by Lieder et al., (2019) reported reduced learning rates in adults with ASD for a contraction bias in pitch discrimination task.

One important limitation among most of these studies is that only the relative effects of priors vs likelihoods are investigated, leaving it unclear – in the case of positive findings – whether the effects originate from altered priors or altered sensory representations. We had addressed this limitation in a recent experiment that allowed us to quantify both priors and likelihoods (Karvelis et al., 2018). We found that autistic traits in the general population were associated decreased biases and that this was underlied not by weaker priors but by increased sensory precision. Here we used the same experimental design to test whether these findings generalize to ASD population.

We studied adults with ASD using a visual motion perception task in which perceptual performance is well characterized by Bayesian inference (Chalk et al., 2010; Karvelis et al., 2018; Valton et al., 2019). In this task, participants have to estimate motion direction of coherently moving cloud of dots (**Fig. 1**). Unbeknownst to participants, two of the motion directions are presented more frequently, which leads to rapid and implicit acquisition of prior expectations for these directions. When ambiguous (low contrast) stimulus is presented, the acquired priors result in attractive estimation biases towards the frequent directions as well as reduced intra-trial variability, reduced reaction times and increased detection performance for the most frequent directions. Finally, when no stimulus is presented, the acquired priors occasionally give rise to hallucinations of the stimulus moving in the expected directions.

Methods

Participants

20 (10 female) individuals with autism spectrum disorder (ASD), and 30 (11 female) controls (CTR) were recruited from advertisements in GP practices and educational settings. The sample size was based on power analysis using the effect size found in Karvelis et al., (2018) and aiming for 80% power with 5% Type I error rate (see Supplementary Material for details). Participants were included if they had normal or corrected-to-normal vision, were able to provide fully informed consent, and had an IQ > 70 (as measured by the Wechsler Abbreviated Scale of Intelligence; Wechsler, 1999). A subsample of 11 participants in the ASD group were interviewed using the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000); where feasible, collateral information was gathered from school reports, clinical letters, and medical notes. Participants with comorbid anxiety or unipolar mood disorders were included. Participants with any neurological disorder, bipolar disorders, attention deficit hyperactivity disorder, psychotic disorders, or current substance use disorder were excluded. All participants were interviewed using the Structured Clinical Interview for DSM-IV (SCID-I; First, Spitzer, Gibbon, & Williams, 2002) to determine inclusion/exclusion criteria. Participants provided informed written consent and were financially compensated for their time and travel. The study received ethical approval from the South East Scotland Research Ethics Committee 01 and NHS Lothian Research & Development.

Apparatus, Stimuli, & Procedure

The experimental setup used in this study was the same as in Karvelis et al. (2018). The stimuli were displayed on a Dell P790 monitor running at 1024 x 768 at 100Hz using Matlab's Psychophysics Toolbox (Brainard, 1997). The visual stimuli consisted of a cloud of dots moving coherently within a circular annulus (see **Fig. 1A**). Four randomly interleaved contrast levels were presented throughout the task: zero contrast (no stimulus; 167 trials), two low contrast levels (90 trials at 2/1 staircase; 243 trials at 4/1 staircase), and one high contrast level (67 trials). The luminance on the high contrast trials was 1.76 cd/m^2 above a 5.18 cd/m^2 background. There were 9 predetermined motion directions in which the stimuli moved: 0º, ±16º, ±32º, ±48º, and

±64º. These directions were set with respect to a central reference angle, which was randomly chosen for each participant. On high contrast trials, 50% of the time the dots moved at completely random directions. The stimuli on low contrast trials moved only at the predetermined motion directions. Across all the non-zero contrasts, the dots moved at ±32º for 58% of the trials, in the other 7 predetermined directions for 36% of the trials and in completely random directions for 6% of the trials. This resulted in a bimodal probability distribution of motion direction (**Fig. 1B**). Participants were not told that stimuli would be presented more at some directions than others.

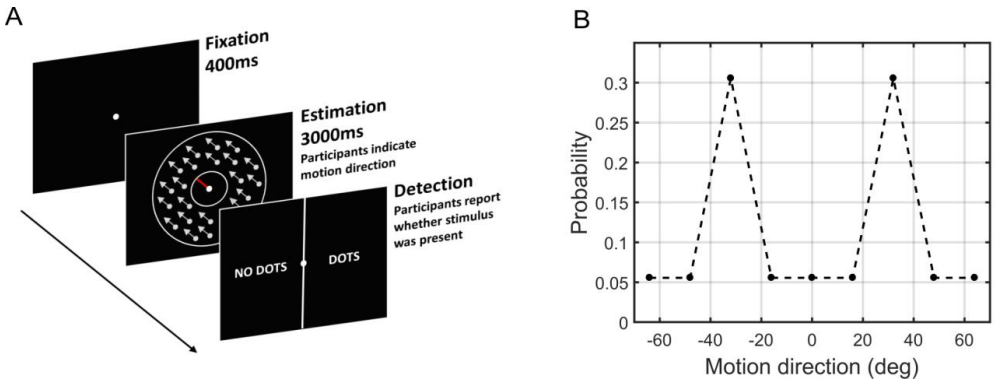


Figure 1: The moving dots task. **(A)** On each trial, participants were presented with a fixation point followed by the motion stimulus and a response bar (red bar). Participants were instructed to align the bar to the perceived motion direction. The screen was cleared either when participants either made an estimation or 3000 ms had elapsed. Lastly, a new screen presented participants with a two-alternative forced choice task (2-AFC), requiring to indicate whether they perceived the dots during the estimation task. **(B)** Probability distribution of the motion directions. Unbeknownst to participants, stimuli moving at ±32° appeared more often than at other motion directions.

On each trial, the participants were first required to estimate the motion direction of dots by aligning the red bar to the direction of motion. Second, in a two-alternative forced choice decision between 'NO DOTS' or 'DOTS', participants were required to indicate whether they detected the stimulus at all by correspondingly clicking on the right or left side of the screen. Feedback on estimation performance was presented every 20 trials by displaying participants' root mean square error, while feedback on detection performance was displayed immediately

after responding by the cursor flashing green or red for correct and incorrect responses, respectively.

Before starting the task, participants were given visual and verbal instructions and completed at least 8 practice trials under supervision of the experimenter, to ensure that the instructions were understood. Participants were instructed to keep their eyes on the fixation point and to respond as quickly and as accurately as they can. The task was completed in a darkened room with a viewing distance of approximately 100cm. Each participant completed 567 trials of the task with breaks every 170 trials. The task took ~45 minutes to complete and participants were fully debriefed at the end.

Behavioural data analysis

Performance on high contrast trials was used as a benchmark to ensure adequate performance in the task. The minimum criteria for inclusion were: 1) detection accuracy of at least 70% and 2) estimation root mean square error (RMSE) of less than 30 degrees. All 30 controls met the criteria, while 2 out of 20 ASD participants did not meet the second criterion and thus were excluded from further data analysis; one more ASD participant was excluded due to very poor detection performance (<30%) on the low contrast trials (**Supplementary Fig. 1)**.

The first 100 trials were excluded from the analysis, as this is how long it took for the staircases to converge to stable contrast levels (**Supplementary Fig. 2**), and for the acquired prior effects to become significant (**Supplementary Fig. 3**). The contrast levels achieved by the 2/1 and 4/1 staircases were found to be overlapping, thus they were combined for all further analysis. Finally, since the distribution of presented directions was symmetrical around a central reference angle, the behavioural measures at equal absolute distance from the reference angle were averaged together to increase the precision of the summary statistics.

The main data analysis was carried out on the estimation performance on low contrast trials (2/1 and 4/1 staircases). Trials on which participants did not make an estimation within the given time (3000 ms) or in the subsequent detection task reported not seeing the stimulus were

excluded from this analysis. To account for any possible lapse estimations when determining bias and variability of responses, the distribution of each individual estimation responses was fitted to the following distribution:

$$(1-\alpha)\cdot V(\mu,\sigma) + {}^{\alpha}/_{2\pi} \qquad (1)$$

where $V(\mu, \sigma)$ is the von Mises circular normal distribution with mean μ and width σ. Parameter α corresponds to the proportion of lapse estimations (uniformly distributed random responses). The estimation bias was calculated as the difference between the estimated mean μ and the true motion direction, while the estimation variability corresponded to σ.

To simultaneously assess the effects of the acquired priors in both groups and to determine whether ASD participants differed from controls, we performed 2-way mixed ANOVA with motion direction as a within-subjects factor and group as a between-subjects factor. We ran Bonferroni-corrected pairwise comparisons to determine which motion directions were driving the effects. All analysis was done using SPSS version 25. To evaluate the strength of the evidence for the null hypothesis we also report Bayes factors (BF_{01}) obtained using the Bayesian statistical software package JASP version 0.10.

Modelling

To control for the possibility of different mechanisms underlying the performance of each individual we fitted a range of models to our data. The first class of models assumed that the biases were of a perceptual nature, as conceived in the Bayesian framework: on every trial, the incoming noisy sensory information is combined with a learned prior of the stimulus statistics, with the mean of the resulting posterior distribution corresponding to the percept (**Fig. 2**). We tested four variants of the Bayesian model (introduced in detail in the next section). Another class of models assumed that task performance could be explained by response strategies that do not involve Bayesian integration (Laquitaine & Gardner, 2018). According to these models, on any given trial participants responded by relying on either the prior or the likelihood alone; the resulting response distribution would thus be a sum (as opposed to a product) of the prior

and the likelihood (hence the class name 'ADD'). We considered four variations of 'ADD' models (see Supplementary Information for details). Below we present the Bayesian models as they provided a better explanation to the data (see **Fig. 4**, model comparison).

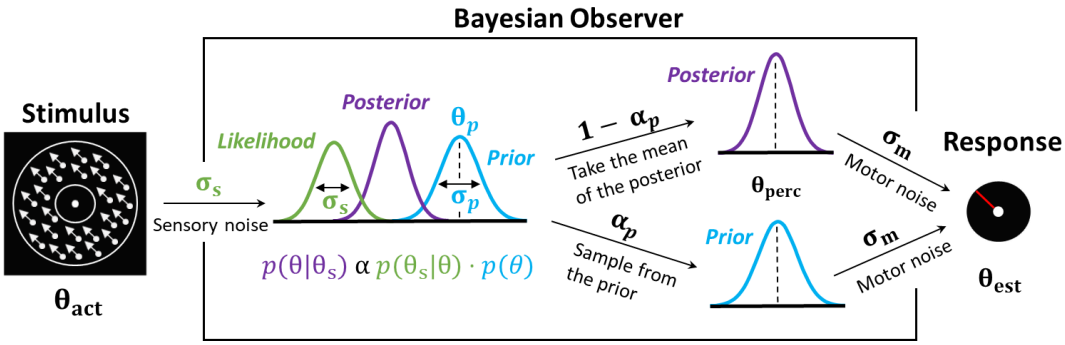


Figure 2. Bayesian model of estimation response for a single trial for the best fitting model (Bayes_P). The actual motion direction (θ_{act}) is corrupted by sensory uncertainty (σ_s), and then combined with prior expectations (mean θ_p and uncertainty σ_p) to form a posterior distribution. The mean of the posterior distribution then corresponds to the perceived motion direction (θ_{perc}). However, on a fraction of trials, determined by the prior-based lapses (α_p), the perceived motion direction is sampled directly from the prior. Finally, in both cases, the response (θ_{est}) is made by perturbing θ_{perc} with motor noise (σ_m). This results in 4 free model parameters: σ_s, σ_p, θ_p and α_p. The motor noise (σ_m) is estimated from high contrast trials and is used as a fixed parameter during the model fitting.

Bayesian models

Following the Bayesian framework, we assumed that participants combined sensory information (likelihood) with their expectations about the motion direction (prior) on every trial. The sensory likelihood of the observed motion direction (θ_s) was parameterised as a von Mises circular normal distribution with variance σ_s:

$$p_{likelihood}(\theta_s|\theta) = V(\theta_s, \sigma_s) \quad (2)$$

The mean of this distribution, (θ_s) depended on the actual presented motion direction (θ_{act}), and to account for trial-to-trial variability it was drawn from another von Mises distribution centred

on θ_{act} with variance σ_s (i.e. V (θ_{act}, σ_s)). Centring the likelihood on the noisy estimate of the actual motion direction reflects the fact that the observer does not have direct access to the actual motion direction and the experimenter cannot know the mean of the internal likelihood an any given trial (Stocker & Simoncelli, 2006; Körding et al, 2007). Since the model is fitted to the distribution of all responses throughout the task (instead of fitting each response individually), such implementation of a varying mean of the likelihood allows us to capture the amount of sensory variability without knowing the exact mean of the internal likelihood on any given trial.

We then hypothesized that participants acquired priors ($p_{prior}(\theta)$) that approximated the bimodal distribution of the stimulus statistics. Therefore, the acquired priors ($p_{prior}(\theta)$) were parameterized as the sum of two von Mises circular normal distributions, centred on motion directions θ_p and $-\theta_p$, each with variance σ_p:

$$p_{prior}(\theta) = \frac{1}{2}[V(-\theta_p, \sigma_p) + V(\theta_p, \sigma_p)] \quad (3)$$

Combining the prior and the likelihood gives us the posterior probability that the stimulus is moving in a direction θ:

$$p_{posterior}(\theta|\theta_s) \propto p_{likelihood}(\theta_s|\theta) \cdot p_{prior}(\theta) \quad (4)$$

The perceived direction, θ_{perc}, was taken to be the mean of the posterior distribution (almost identical results would be obtained by using the maximum instead).

Finally, we accounted for motor noise (i.e. inaccuracies when aligning and clicking the mouse) and lapse estimations, such that:

$$p(\theta_{est}|\theta_{perc}) = (1 - \alpha_p) \cdot V(\theta_{perc}, \sigma_m) + \alpha_p \cdot [p_{prior}(\theta) * V(0, \sigma_m)] \quad (5)$$

where the asterisk ($*$) denotes convolution, σ_m is the motor noise and α_p is the proportion of prior-based lapse estimations on each trial (i.e. lapse estimations that follow the participants'

acquired expectations – $p_{prior}(\theta)$). We called this model 'BAYES_P' for Bayes with Prior-based lapses.

We also tested a simpler variant of this model which assumed that the lapse estimations (Eq. (5)) were not made based to the acquired prior but instead were completely random (model 'BAYES'). Furthermore, to account for the possibility that in addition to the acquisition of the prior there might also be adaptations in the sensory likelihood itself (e.g.,Sato and Körding, 2014), we tested two other variants of this model: 'BAYES_var' where the sensory precision varied with the stimulus direction (i.e. it took five different values for each of the angles: 0°, ±16°, ±32°, ±48°, ±64°) and 'BAYES_varmin' where sensory precision was allowed to be different for ±32° but was the same for all other directions. BAYES_P and BAYES had a total of 4 free parameters, while BAYES_varmin and BAYES_var had 5 and 8, respectively.

Note that in our previous work on ASD traits (Karvelis et al., 2018), we did not test BAYES_P model and our winning model was BAYES. However, the results of that study – a negative correlation between AQ and sensory uncertainty – remains unchanged when BAYES_P model is fit to the data (**Supplementary Fig. 6**). Furthermore, ADDr family of models presented here is also slightly different from that used in Karvelis et al. (2018). These models still implement the same ideas but following recent work by Laquitaine & Gardner (2018) they were parameterized more parsimoniously (hence the added letter 'r' for 'reduced'), rendering the comparison with the Bayesian models fairer (see Supplementary Material for details)

Results

Participant characteristics

A predefined performance criterion of at least 70% detection and a root mean square error of 30° or less on high contrast trials was used as an indicator of adequate task performance. Two ASD participants scored below this criterion (**Supplementary Fig. 1**) and were therefore excluded from all subsequent analyses. One more ASD participant was excluded due to exceptionally poor performance on low contrast trials (detection < 30%). The characteristics of the included participants are summarised in Table 1.

Table 1. *Participant Characteristics (standard deviation in parentheses)*

	ASD (n = 17)	CTR (n = 30)	*Statistic*	*p*
Age (years)	32.00 (13.87)	34.52 (11.12)	Z = 1.08	.28
Gender (M:F)	7:10	19:11	χ^2 = 2.15	.14
Full-scale IQ	119.18 (9.07)	118.38 (7.21)	Z = -0.32	.75
Performance IQ	116.44 (12.07)	118.35 (9.55)	Z = 0.09	.93
Verbal IQ	116.33 (11.25)	114.65 (9.13)	Z = -0.61	.54
GAF	58.53 (9.35)	74.79 (10.70)	Z = 4.23	**<.001**
AQ	36.35 (6.50)	13.37 (8.86)	Z = -4.79	**<.001**
ASRS	40.71 (11.09)	29.79 (10.49)	Z = -2.62	**.01**

Note. ASD = Autism-Spectrum Disorders; CTR = Controls; GAF = Global Assessment of Functioning (potential range 1-100; higher is better); AQ = Autism-Spectrum Quotient (potential range 0-50); ASRS = Adult ADHD Self-Report Scale (motential range 0-72). Group comparisons were done using Wilcoxon rank-sum test and Chi square test (for gender).

Detection performance and contrast levels

To ensure that the stimulus was equally ambiguous for each participant despite the individual differences in contrast sensitivity, the stimulus contrast adjusted based on the detection performance during the task. Using 2/1 and 4/1 staircases, we ensured that the individual detection performances would converge to 70.4% and 84.1% respectively (Levitt, 1971).

Contrast staircases converged to stable luminance levels after about 100 trials for both groups and the reached levels did not differ between the groups (**Supplementary Fig. 2**). Controls converged to 0.38 cd/m^2 (±0.03) for 2/1 staircase and 0.41 cd/m^2 (±0.03) for 4/1 staircase while patients converged to 0.38 cd/m^2 (±0.03) for 2/1 staircase and 0.43 cd/m^2 (±0.03) for 4/1 staircase.

Statistical learning

First, we investigated whether participants acquired the statistics of the stimulus. We looked at measures suggestive of statistical learning in each group, namely: attractive biases towards the most frequent directions, decreased variability, decreased reaction times, improved detection performance for the most frequent directions and hallucinations on trials with no stimulus (**Fig. 3**).

Estimation performance

To investigate whether the participants' perceived motion directions were biased, we measured the difference between the presented and reported motion directions. **Fig. 3A** displays the average estimation bias plotted against the presented motion direction for each group. Overall, there was a significant effect of motion direction on the estimation bias ($F(2.51, 191.60) = 12.62$, $p < 0.001$, $\eta_p^2 = 0.219$, Greenhouse-Geisser correction $\varepsilon = 0.628$), but no differences between the groups (group main effect: $F(1, 45) = 0.44$, $p = 0.511$, $\eta_p^2 = 0.010$; $BF_{01} = 4.60$); and no group*angle interaction ($F(2.51, 191.60) = 1.12$, $p = 0.342$, $\eta_p^2 = 0.024$). Pairwise comparisons (with Bonferroni correction) revealed that there was an attractive bias towards ±32° at ±64° (MD = 13.45, $p < 0.001$) and at ±48° (MD = 7.80, $p < 0.001$). Together, these results suggest that both groups were similarly biased towards most frequently presented directions, consistent with having acquired the priors that approximate the statistics of the stimulus.

Next, we analysed how acquired priors affected variability of motion estimations. (**Fig. 3B**). We found a significant main effect of motion direction ($F(2.95, 132.83) = 8.80$, $p < 0.001$, $\eta_p^2 = 0.164$,Greenhouse-Geisser correction $\varepsilon = 0.738$), but no conclusive group effects (main effect of group: $F(1, 345) = 3.85$, $p = 0.056$, $\eta_p^2 = 0.079$; $BF_{01} = 0.80$); and no group*angle interaction $F(2.95, 132.83) = 1.23$, $p = 0.302$, $\eta_p^2 = 0.027$). Pairwise comparisons (with Bonferroni correction) revealed that the main effects of motion direction were driven by the variability at ±32° being lower than at 0° (MD = 7.08, $p < 0.001$), at ±64° (MD = 4.60, $p = 0.001$) and at ±16° (MD = 3.48, $p = 0.02$). This result is also consistent with the participants having acquired the priors that approximate the statistics of the stimulus, but points to possible differences between groups in overall perceptual variability

Reaction times and detection

We also examined whether participants' acquired expectations reduced the reaction times at the expected motion directions **(Fig. 3D**). There was a significant main effect of motion direction on estimation reaction ($F(3.58, 161.16) = 10.11$, $p < 0.001$ $\eta_p^2 = 0.184$, Huynh-Feldt correction $\varepsilon = 0.895$). This was driven by decreased reaction times at the most frequent directions as revealed by pairwise comparisons: reaction time at ±32° was significantly shorter than at 0° (MD = 0.86, p = 0.005) at ±64° (MD = 0.12, p < 0.001) and at ±48° (MD = 0.45, p = 0.044). Furthermore, ASD participants were significantly faster than controls ($F(1, 45) = 8.30$, $p = 0.006$, $\eta_p^2 = 0.156$), but there was no interaction between group and motion direction ($F(3.58, 161.16) = 1.78$, $p = 0.141$, $\eta_p^2 = 0.038$).

We then analysed whether the acquired expectations improved detection at the expected motion directions (**Fig. 3E**). The detection of stimulus had a significant main effect of motion direction ($F(2.38, 107.28) = 7.69$, $p < 0.001$, $\eta_p^2 = 0.146$, Greenhouse-Geisser correction $\varepsilon = 0.596$), with stimulus at ±32° being the most frequently detected direction as shown by pairwise comparisons: detection at ±32° was significantly better than at all other directions (0°, MD = 7.71, p = 0.004; ±16°, MD = 4.65, p = 0.003; ±48°, MD = 6.75, p < 0.001; ±64°, MD = 9.43, p < 0.001). The main group effect for detection could not be present because using 1/2 and 1/4 contrast staircases guarantees that everyone has the same average detection rate, while group*motion direction interaction was non-significant: $F(2.38, 107.28) = 1.43$, $p = 0.243$ $\eta_p^2 = 0.031$).

Hallucinations

Finally, we investigated whether the acquired priors of the motion directions affected the participants' perception on trials where no stimulus was presented, but where participants reported both a motion direction (in estimation part) and seeing a stimulus (in detection part). We refer to this effect as hallucinations. To quantify whether participants were more likely to hallucinate motion stimulus around the most frequently presented motion directions, we multiplied the probability that participants estimated within 16° of these motion directions by the total number of bins:

$$p_{ratio} = p(\theta_{est} = \pm 32(\pm 16)°) \cdot N_{bins} \qquad (6)$$

This probability ratio would be equal to 1 if participants were equally likely to estimate within 16° of ±32° as they were to estimate within the other bins of equal size.

We found that the median value of *'p_{ratio}'* was significantly greater than 1 for both ASD and control participants (median(p_{ratio}) = 2.92, p = 0.001 and median(p_{ratio}) = 3.00, p < 0.001, respectively; two-tailed signed-rank test), indicating that both groups hallucinated significantly more around the most frequent motion directions as opposed to all other directions (**Fig. 3F**). Bayesian statistical analysis provided anecdotal evidence for the groups being the same in this measure (BF_{01} = 2.52). Next, we also tested whether ASD participants differed from controls in the number of total hallucinations experienced in the task, but no differences were found (p = 0.535, two-tailed rank-sum test; BF_{01} = 2.93).

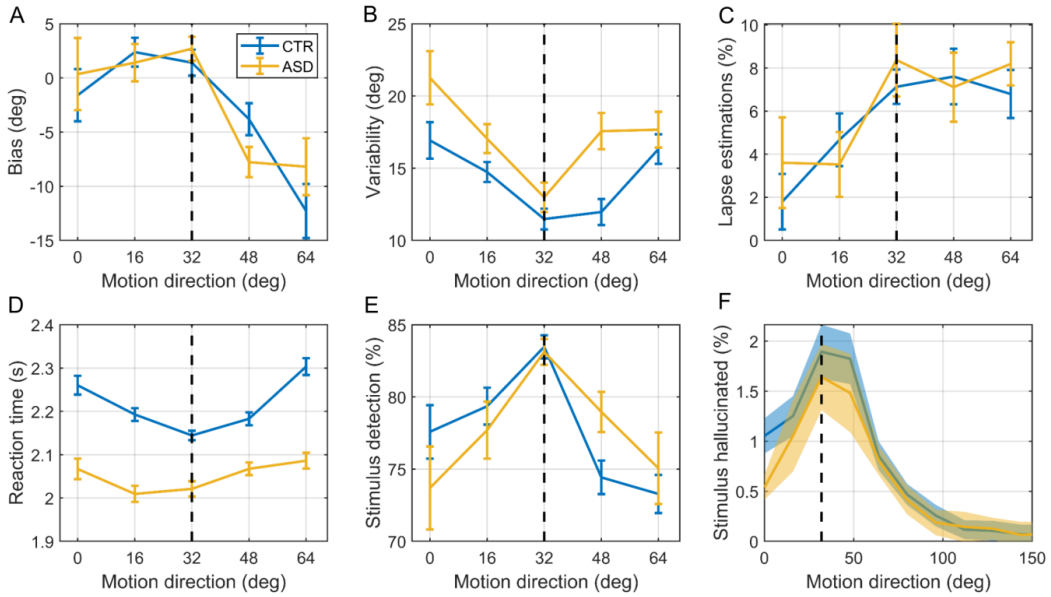


Figure 3. Performance on **(A-E)** low contrast trials and **(F)** no stimulus trials by control (blue lines) and ASD (yellow lines) participants. (**A**) Mean estimation bias (**B**) estimation standard deviation (i.e. variability) (**C**) lapse estimations, (**D**) reaction times during the estimation task, (**E**) the fraction of trials in which the stimulus was detected, (**F**) the fraction of no stimulus trials in which the stimulus was hallucinated. The error bars and shaded areas represent within-subject standard error. The vertical dashed lines correspond to the most frequently presented motion directions (i.e. ±32°).

ADOS, AQ and behaviour

Next, we performed an exploratory analysis of the relationship between ASD severity and the main measures of interest: bias, variability and hallucinations. While all of participants in the ASD group had a previous diagnosis, not all of them scored above threshold when assessed with ADOS at the time of testing. We investigated whether ADOS positive vs ADOS negative group assignment was associated with performance differences and we also performed correlational analysis using total ADOS scores. Together, these results showed substantial evidence for ADOS having no effect on bias and variability, while for hallucinations the results remained inconclusive (**Supplementary Fig. 4**).

We also investigated whether the negative relationship between AQ and the measures of bias, variability and hallucinations reported in Karvelis et al., (2018) was present in this sample. We examined this in each group separately, as well as pooling the data across the groups. The results converged to a substantial evidence for the null hypothesis for bias and variability, while for hallucinations it remained inconclusive (**Supplementary Fig. 5A-C**).

Modelling results

We fitted the models to the behavioral data and computed the Bayesian Information Criterion (BIC). For robustness, we performed both fixed effects model comparison (summed BIC across individuals for each model) and random effects Bayesian model selection. Both analyses selected BAYES_P model (**Fig. 2**) as the winning model for both groups (**Fig. 4**). Model fits to the data also showed that BAYES_P was much better at fitting lapse estimations, confirming that such estimations followed the acquired prior distribution instead of being random (**Fig. 5C and G**).

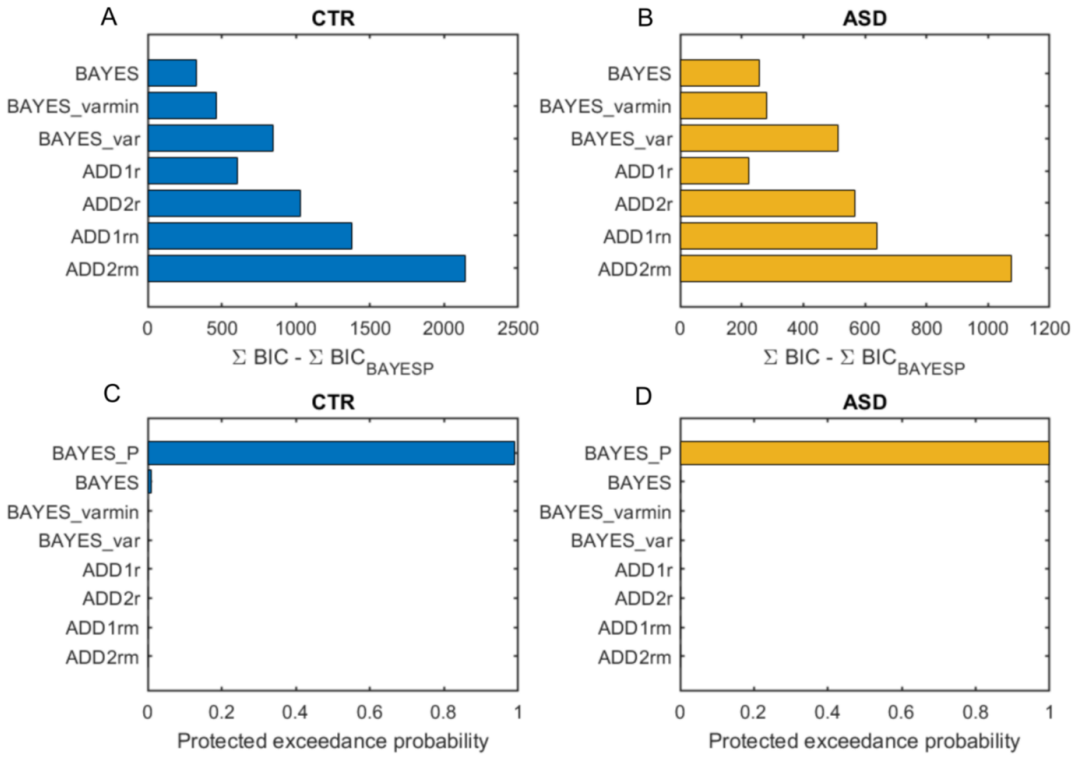


Figure 4. Model comparison and selection. (**A, B**) Fixed effects model selection using Bayesian Information Criterion (BIC). X-axis measures the relative difference between BIC of each model (as indicated on Y-axis) and BIC of BAYES_P (winning model) summed across participants. Smaller BIC indicate a better model. For both groups BAYES_P provided the best model evidence. (**C, D**) Random effect Bayesian model selection. Higher protected exceedance probability indicates a model having a higher likelihood of being more frequent among the subjects. For both groups BAYES_P was the most likely model.

Finally, we compared patients with controls on the basis of BAYES_P parameters (**Fig. 4I-L**). Consistent with the behavioral data results, none of the parameters were found to be different between the groups: the mode of the prior ($p = 0.213$, $BF_{01} = 1.71$) precision of the prior ($p = 0.939$; $BF_{01} = 3.28$), precision of sensory likelihood ($p = 0.197$, $BF_{01} = 2.10$) and prior-based lapse estimations ($p = 0.974$, $BF_{01} = 3.63$). However, ASD group was less variable in the mode of the learnt prior ($F(29,16) = 3.28$, $p = 0.015$).

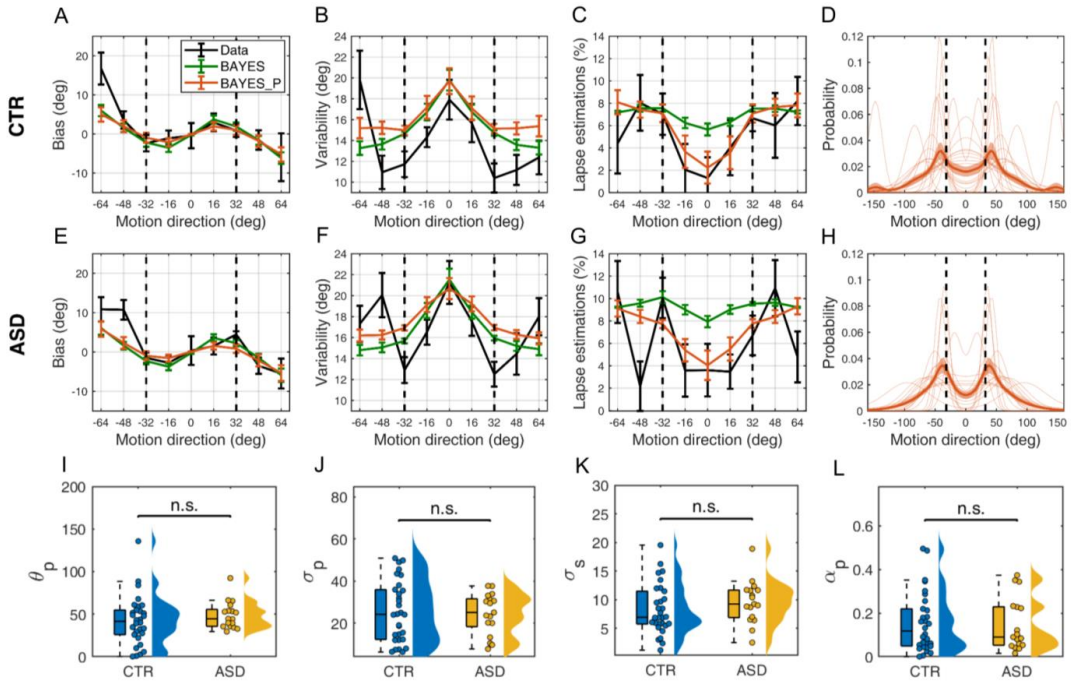


Figure 5. Model fits and parameter estimates. **(A-H)** Model fits for the best fitting model BAYES_P (orange) and the second-best model BAYES (green), to the behavioral data (black). **(A-D)** Controls and **(E-H)** ASD participants. **(A, E)** Estimation bias, **(B, F)** estimation variability, **(C, G)** estimation lapse rate, **(D, H)** prior expectations of each individual (thin orange lines) and group average (thick orange line) as estimated via BAYES_P model. The vertical dashed lines correspond to the most frequently presented motion directions (i.e. ±32°). The error bars and shaded areas represent within-subject standard error. **(I-L)** Comparison of BAYES_P model parameter values for controls and ASD participants; jittered dots denote individual participants, colored areas represent density of the data points. **(I)** θ_p – the mean of acquired prior ($p = 0.213$, $BF_{01} = 1.71$), **(J)** σ_p – the uncertainty in the acquired prior ($p = 0.939$; $BF_{01} = 3.28$), **(K)** σ_s – the uncertainty of sensory likelihood ($p = 0.197$, $BF_{01} = 2.10$), **(L)** α_p – prior-based lapse rate ($p = 0.974$, $BF_{01} = 3.63$). n.s. – non-significant.

Discussion

This study was motivated by the hypothesis that ASD is associated with a systematic imbalance in uncertainty inherent in sensory evidence ('likelihood') and prior expectations ('prior') (Pellicano & Burr, 2012; Brock, 2012; Lawson et al., 2014; Van de Cruys et al., 2014). More specifically, we sought to investigate whether the proposed imbalance arises from the likelihood or the prior, and if it is the latter, how it relates to rapid prior acquisition. We used a visual

motion estimation task (Chalk et al., 2010) in which priors are induced implicitly by presenting two of the motion directions more frequently. Importantly, the design of this task and the use of computational modelling enabled us to disentangle and quantitatively assess both priors and likelihoods.

We found that ASD group showed no deficits in acquiring motion direction priors and exhibited comparable estimation biases and variability. Modelling results confirmed that ASD group did not exhibit any imbalances in weighting priors and likelihoods. At first glance, this stands in contradiction to our previous work using this same experimental design in the general population where we found that autistic traits - measured with Autism-Spectrum Quotient (AQ) - were associated with more precise sensory likelihoods, which was expressed in reduced estimation biases and variability (Karvelis et al., 2018).

One factor to consider for explaining such inconsistency would be our ASD sample. While all of participants in the ASD group had a previous diagnosis, it is worth noting that only 4 out of 11 assessed with ADOS at the time of testing scored above the threshold. This raises a possibility that our negative result might be due to the ASD sample being unrepresentative of the ASD population. However, an exploratory analysis of ADOS and performance did not show any clear trends to support such idea (**Supplementary Fig. 4**).

Another factor to consider for explaining the null finding is the assumption of continuity between autistic traits in the general population and ASD diagnosis. Using autistic traits to investigate hypotheses about ASD population is widespread and is justified by multiple studies reporting consistent effects along autistic traits and in clinical ASD: in abnormal sensory experiences (Horder et al., 2014; Robertson & Simmons, 2013) and in a variety of experimental tasks such as global dot motion (Grinter et al., 2009), visual search (Grinter et al., 2009; Almeida et al. 2010; Brock et al. 2011), block design test (Stewart et al. 2009) and emotion recognition (Baron-Cohen et al., 2001). While it is not surprising that findings in clinical ASD might not always generalize to autistic traits in neurotypicals (e.g., due to diminishing effects), it is much more difficult to explain why effects present along autistic traits would not be seen in the ASD population. Nonetheless, our study is not the only one in this regard. For instance, Ewbank et al. (2014, 2017) found reduced repetition suppression along autistic traits for faces, scenes and

simple shapes, while in a clinical ASD sample the effect was found only for faces but not shapes. A recent study by Lawson et al. (2018) investigating adaptation to social and non-social stimuli found autistic traits to be associated with reduced adaptation across both stimulus types, while adults with ASD diagnosis showed reduced adaptation only for the social eye-gaze stimulus. A possible explanation might be that factors other than the autistic endophenotype – such as generalized anxiety disorder (Ashwood et al., 2016) – contribute to autistic traits scores and are responsible for the observed inconsistencies (e.g., see Gregory & Plaisted-Grant, 2016, for discussion). Thus, while plenty of evidence supports continuum between autistic traits in neurotypicals and clinical ASD, caution should be taken when drawing conclusions about ASD based on the effects found along autistic traits.

It is also worth noting that in the current controls group we did not replicate previous findings (Karvelis et al., 2018) of the negative relationship between AQ and sensory uncertainty as estimated via model fitting, nor with the associated behavioural measures of bias, variability and hallucinations (**Supplementary Fig. 5**). Pooling the data across the groups also showed evidence against the negative relationship between AQ and sensory uncertainty and the associated behavioural measures (**Supplementary Fig. 5**).

Our findings of intact acquisition of motion direction priors, together with mixed findings of previous studies (intact priors: Spanò et al., 2016, Pell et al., 2016; Croydon et al., 2017; Van de Cruys et al., 2018; Utzerath et al., 2019; weaker priors: Powell et al., 2016; Skewes et al., 2015; Karaminis et al., 2016), speak against the idea that reduced reliance on prior in ASD is a general impairment. Further evidence against this idea comes from studies investigating individual differences in prior influence across different tasks and stimuli. Tulver et al. (2019) assessed prior effects in four different tasks within the same sample of participants: Kanizsa's square illusion, detection of blurred words/non-words, two-tone (Mooney) face recognition and representational momentum task. Autistic traits were not found to modulate the strength of prior effects in any of the tasks. Furthermore, while correlations were found between some of the tasks, no common factor was found to explain the variability in prior effects across all tasks. Even within the domain of perceptual illusions alone such factor (whether mediated via autistic phenotype or not) does not show any signs of existing. Chouinard et al. (2016) used a battery of 13 perceptual illusions and found reduced susceptibility along autistic traits to only 2 (Shepard's

table-tops and Square-diamond). Similarly, Grzeczkowski et al. (2017, 2018) used a battery of 7 perceptual illusions and found no correlations in illusion magnitude among most of the illusions, again suggesting no common factor for reliance on priors. All in all, these findings suggest that individual and ASD population differences in 'reliance on priors' do not generalize across different tasks and stimuli, even when the latter are very similar (e.g., in the case of similar perceptual illusions).

More recent and more refined versions of the Bayesian accounts of ASD move beyond the simple idea of chronic prior and sensory precision imbalance and suggest this imbalance to be dependent on the hierarchical organization of environmental features and their (in)stability over time, i.e. volatility (for a review see Palmer et al., 2017). This would reformulate the hypothesis in terms of aberrant context-sensitive precision modulation in ASD. While no model-based experimental studies of hierarchical inference of environmental features have yet been reported, a couple of studies have deployed Hierarchical Gaussian Filter (HGF) (Mathys et al., 2014) to study inference and learning in ASD in changing environments. Sevgi et al. (2016) used HGF to analyse learning along non-clinical autistic traits in a reward-based learning task with social and non-social cues, reliability of which varied throughout the task. No differences in the hierarchical estimation of volatility was found, however autistic traits were associated with decreased reliance on the social cue at the decision-making stage. Lawson et al. (2017) used HGF to study adults with ASD in audio-visual associative learning task. ASD group was found to have higher meta-volatility estimates, which lead to smaller adjustments of their learning rate about stimuli contingencies when transitioning from stable to volatile phases of the task, while on average the learning rates were not found to be different.

Another important step for ASD models in addition to incorporating the hierarchy of features and volatility would be to include the coupling between perception and action, i.e., active inference - this would account for differences in how information is sampled, which could underlie different learning trajectories and different inferences. While plenty of findings exist to motivate the development of models that integrate all of these elements (for a discussion see Palmer et al., 2017), they are yet to be developed and tested in the ASD population.

References

Almeida, R. A., Dickinson, J. E., Maybery, M. T., Badcock, J. C., & Badcock, D. R. (2010). Visual search performance in the autism spectrum II: The radial frequency search task with additional segmentation cues. *Neuropsychologia*, *48*(14), 4117-4124.

Ashwood, K. L., Gillan, N., Horder, J., Hayward, H., Woodhouse, E., McEwen, F. S., ... & Cadman, T. (2016). Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. *Psychological medicine*, *46*(12), 2595-2604.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. The Journal of Child Psychology and Psychiatry and Allied Disciplines, 42(2), 241-251.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433-436.

Brock, J., Xu, J. Y., & Brooks, K. R. (2011). Individual differences in visual search: Relationship to autistic traits, discrimination thresholds, and speed of processing. *Perception*, *40*(6), 739-742.

Chalk, M., Seitz, A.R., Seriès, P., (2010). Rapidly learned stimulus expectations alter perception of motion. Journal of Vision 10, 2.

Ewbank, M. P., Rhodes, G., von dem Hagen, E. A., Powell, T. E., Bright, N., Stoyanova, R. S., Baron-Cohen, S., & Calder, A. J. (2014). Repetition suppression in ventral visual cortex is diminished as a function of increasing autistic traits. *Cerebral Cortex*, *25*(10), 3381-3393.

Ewbank, M. P., Pell, P. J., Powell, T. E., von dem Hagen, E. A., Baron-Cohen, S., & Calder, A. J. (2017). Repetition suppression and memory for faces is reduced in adults with autism spectrum conditions. Cerebral Cortex, 27(1), 92-103.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (2002). *Structured clinical interview for DSM-IV-TR axis I disorders, research version*. New York, NY: American Psychiatric Publishing.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. Trends in cognitive sciences, 14(3), 119-130.

Friston, Karl. 2005. "A Theory of Cortical Responses." Philosophical transactions of the Royal Society of London. Series B, Biological sciences 360(1456): 815–36.

Gregory, B. L., & Plaisted-Grant, K. C. (2016). The Autism-Spectrum Quotient and

Visual Search: Shallow and Deep Autistic Endophenotypes. Journal of Autism and Developmental Disorders, 46(5), 1503–1512.

Grinter, E. J., Maybery, M. T., Van Beek, P. L., Pellicano, E., Badcock, J. C., & Badcock, D. R. (2009). Global visual processing and self-rated autistic-like traits. *Journal of Autism and Developmental Disorders*, *39*(9), 1278-1290.

Grzeczkowski, L., Clarke, A. M., Francis, G., Mast, F. W., & Herzog, M. H. (2017). About individual differences in vision. Vision research, 141, 282-292.

Grzeczkowski, L., Roinishvili, M., Chkonia, E., Brand, A., Mast, F. W., Herzog, M. H., & Shaqiri, A. (2018). Is the perception of illusions abnormal in schizophrenia?. Psychiatry research, 270, 929-939.

Horder, J., Wilson, C. E., Mendez, M. A., & Murphy, D. G. (2014). Autistic traits and abnormal sensory experiences in adults. *Journal of autism and developmental disorders*, *44*(6), 1461-1469.

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, 28570.

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.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. TRENDS in Neurosciences, 27(12), 712-719.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. PLoS one, 2(9), e943.

Król, M., & Król, M. (2019). The world as we know it and the world as it is: Eye-movement patterns reveal decreased use of prior knowledge in individuals with autism. Autism Research.

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. The Journal of the Acoustical society of America, 49(2B), 467-477.

Lieder, I., Adam, V., Frenkel, O., Jaffe-Dax, S., Sahani, M., & Ahissar, M. (2019). Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. Nature neuroscience, 22(2), 256-264.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., ... & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of autism and developmental disorders, 30(3), 205-223.

Mathys, C.D. et al. Uncertainty in perception and the Hierarchical Gaussian Filter. Front. Hum. Neurosci. 8, 825 (2014).

Palmer CJ, Lawson RP, Hohwy J. 2017. Bayesian approaches to autism: Towards volatility, action, and behavior. Psychological Bulletin 143:521–542

Rao, R. P. N. & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2, 79–87.

Robertson, A. E., & Simmons, D. R. (2013). The relationship between sensory sensitivity and autistic traits in the general population. *Journal of Autism and Developmental disorders*, *43*(4), 775-784.

Sato, Y., & Körding, K. P. (2014). How much to trust the senses: Likelihood learning. Journal of Vision, 14(13), 13-13.

Seriès, P., & Seitz, A. (2013). Learning what to expect (in visual perception). Frontiers in human neuroscience, 7, 668.

Sevgi, M., Diaconescu, A. O., Tittgemeyer, M., & Schilbach, L. (2016). Social Bayes: Using Bayesian Modeling to Study Autistic Trait–Related Differences in Social Cognition. Biological Psychiatry, 80(2), 112–119.

Skewes, J. C., Jegindø, E. M., & Gebauer, L. (2015). Perceptual inference and autistic traits. Autism, 19(3), 301-307.

Stewart, M. E., Watson, J., Allcock, A. J., & Yaqoob, T. (2009). Autistic traits predict performance on the block design. *Autism*, *13*(2), 133-142.

Spanò, G., Peterson, M. A., Nadel, L., Rhoads, C., & Edgin, J. O. (2016). Seeing can be remembering: Interactions between memory and perception in typical and atypical development. Clinical Psychological Science, 4(2), 254–271.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. Nature neuroscience, 9(4), 578.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. Psychological review, 121(4), 649.

Valton, V., Karvelis, P., Richards, K. L., Seitz, A. R., Lawrie, S. M., & Seriès, P. (2019). Acquisition of visual priors and induced hallucinations in chronic schizophrenia. *Brain*, *142*(8), 2523-2537.

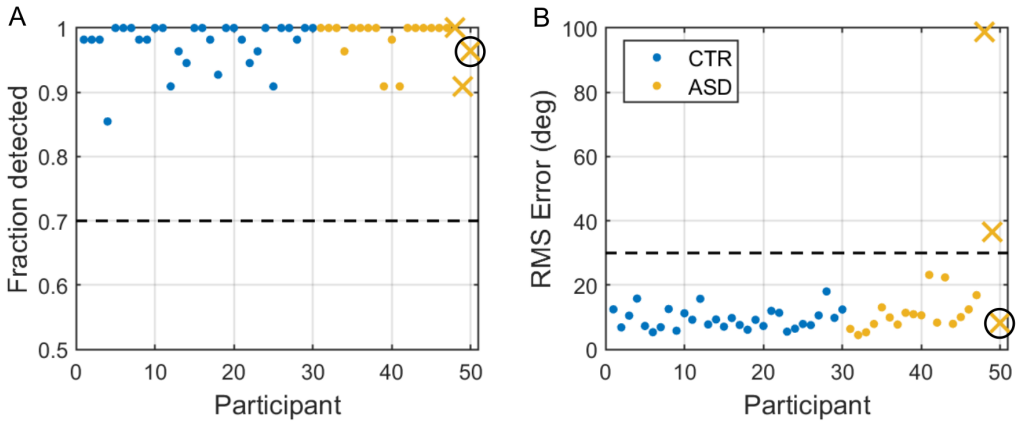
Wechsler, D. (1999). *Manual for the Wechsler abbreviated intelligence scale (WASI)*. San Antonio, TX: The Psychological Corporation.

Supplementary Material

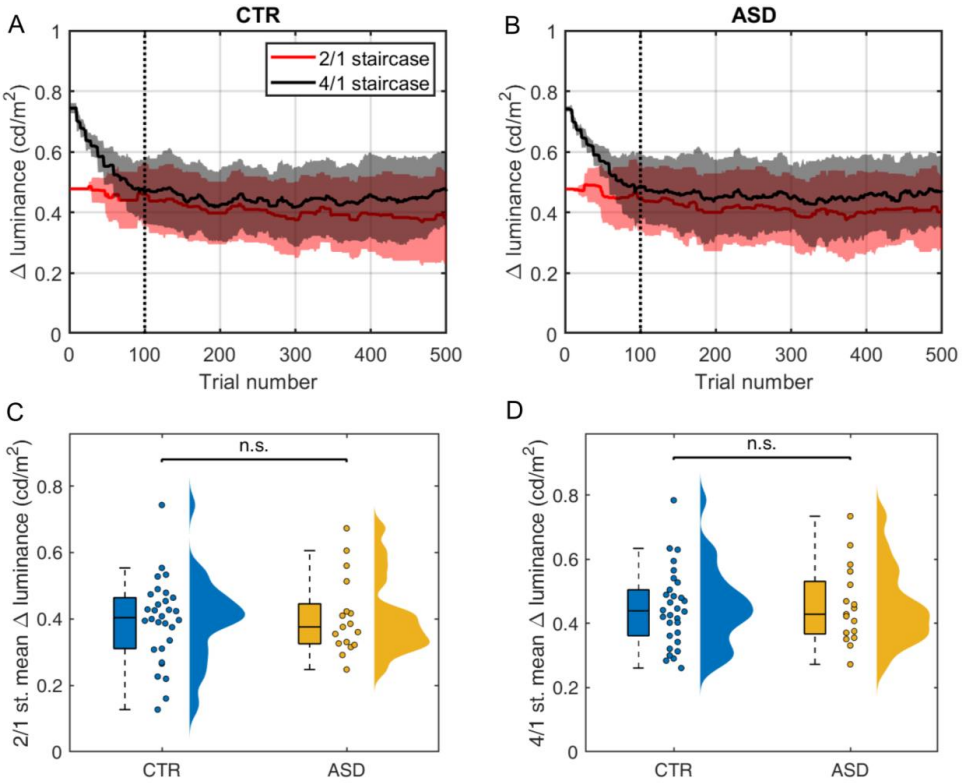
Power analysis

To determine the minimal sample size, we performed an a-priori power analysis. To estimate the effect size we would expect to find in the ASD population, we relied on our previous findings on ASD traits in the general population using the same task design (Karvelis et al., 2018). To obtain Cohen's d from the correlational effects found in Karvelis et al. (2018), we dichotomized the data along Autism-Spectrum Quotient (AQ) and compared ~25% lowest scores with ~25% highest scores: 20 participants within 0-15 AQ score vs. 23 participants within 25-41 AQ score, respectively. This resulted in the group AQ means being different by ~20 points, which is also the AQ difference found between the general population and ASD population (Ruzich et al., 2015). We estimated effect sizes for two measures: 1) estimation variability, which was the strongest effect expressed behaviourally, and 2) sensory precision, which was derived from fitting a generative Bayesian model. For estimation variability Cohen's d was found to be 1.01, while for sensory precision it was 0.81.

We calculated that for 80% power and type I error rate of 5%, 14 participants in each group would suffice to detect behavioural effects in estimation variability, while for sensory precision we would need 21 participants per group, or alternatively, 16 ASD participants for a sample of 30 controls. These calculations were done using GPower 3 software (Faul et al., 2007).



Supplementary Figure 1. Performance on high contrast trials: (**A**) detection rate and (**B**) rms error (right). Dashed lines indicate the exclusion criteria (70% detection and 30° RMS error). Participants who did not satisfy at least one of these criteria are denoted with a cross marker and were excluded from further analysis. In addition, one more participant was excluded (circled) due to poor detection performance on low stimulus trials (<30%). This resulted in 30 controls (CTR), 17 autism spectrum disorder (ASD) participants suitable for further data analysis.

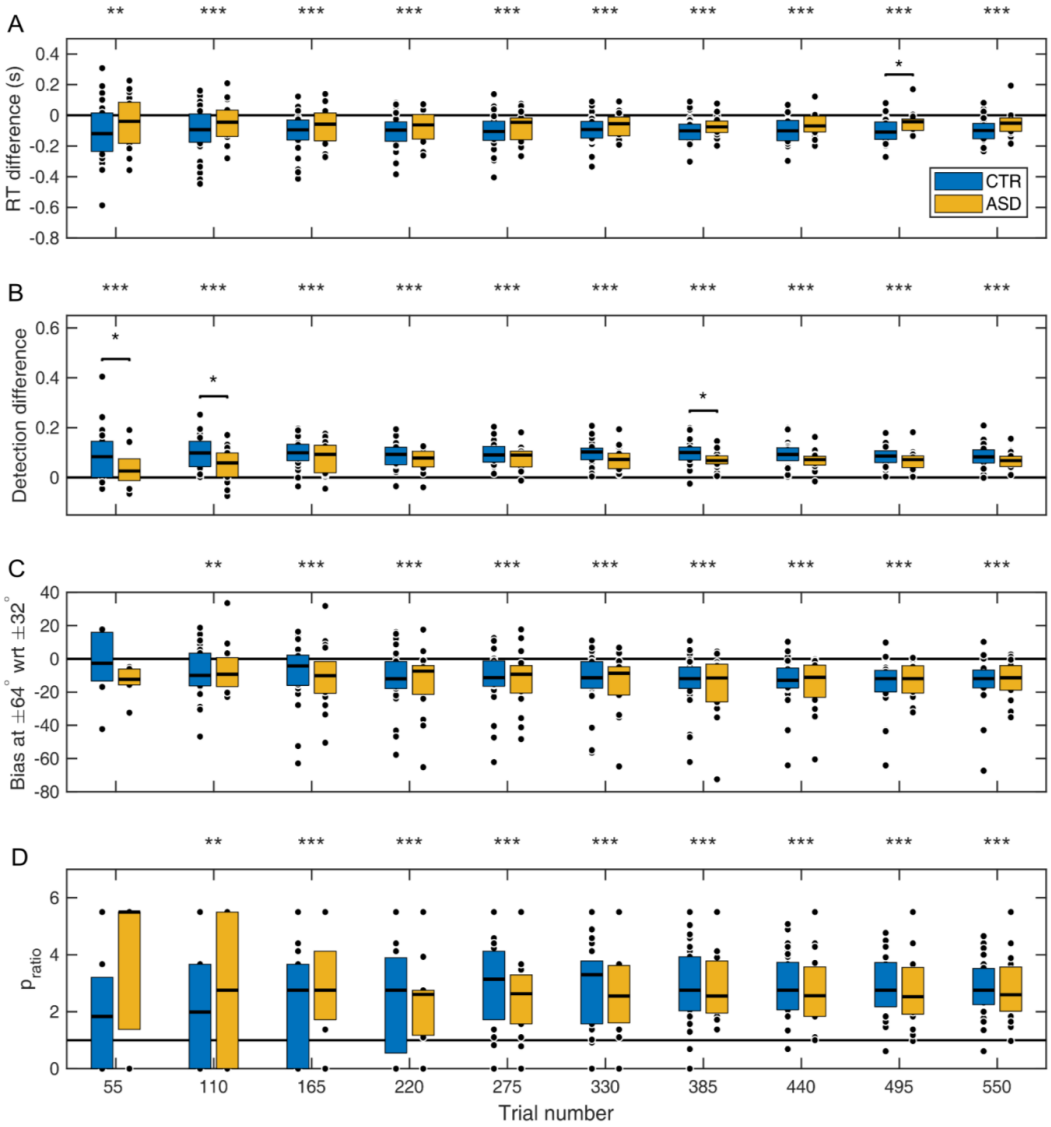


Supplementary Figure 2. 2/1 and 4/1 staircase contrast levels throughout the task. (A) Controls, (B) ASD participants. Both groups were found to reach convergence contrast levels after around 100 trials. These trials were removed from further data analysis The groups did not differ in the achieved contrast levels (C, D). n.s. – non-significant

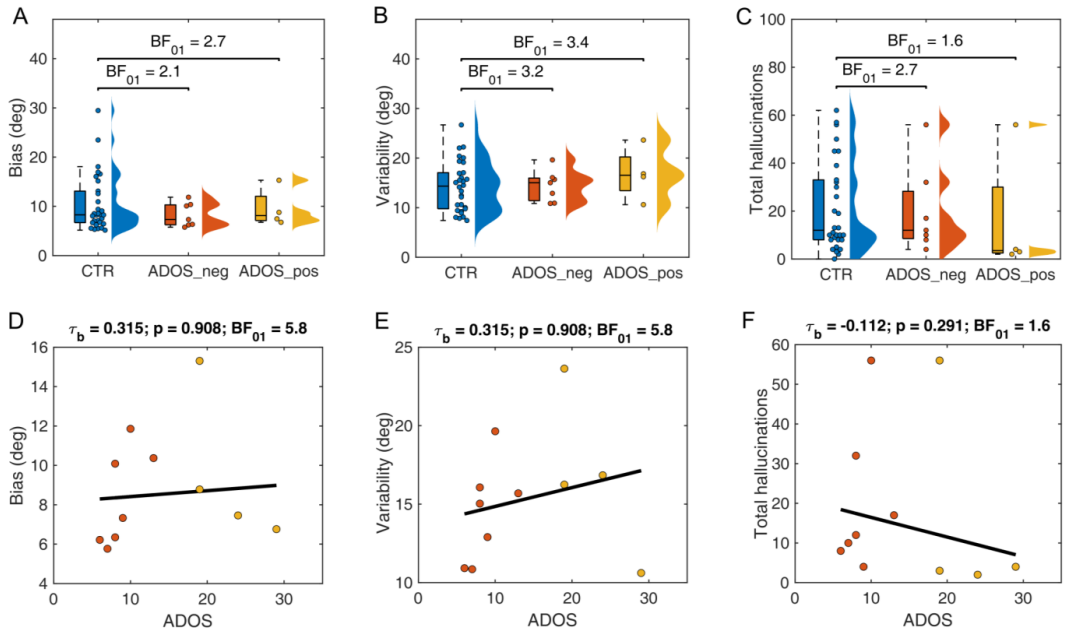
Emergence of prior effects

We investigated at which point in the task the acquired expectations started to have a significant effect on performance and whether this was different for patients. To do so, we computed cumulative moving averages at every 55 trials for different measures of interest and tested for significance of the prior effects (**Supplementary Fig. 3**). In the order presented, we tested when the estimation reaction times (RT) at ±32° became shorter than at all other directions, when the detection at ±32° became higher than at all other directions, when the bias at ±64° become more negative thenbias at ±32° and when the probability of hallucinating within 16° of ±32° became larger than at other directions (p_ratio; Eq. (6)). For all of these measures we performed one-tailed Wilcoxon signed rank test pooling data across the groups to test for the effects of the prior, and two-tailed Wilcoxon rank-sum test for comparing the groups at each step of 55 trials.

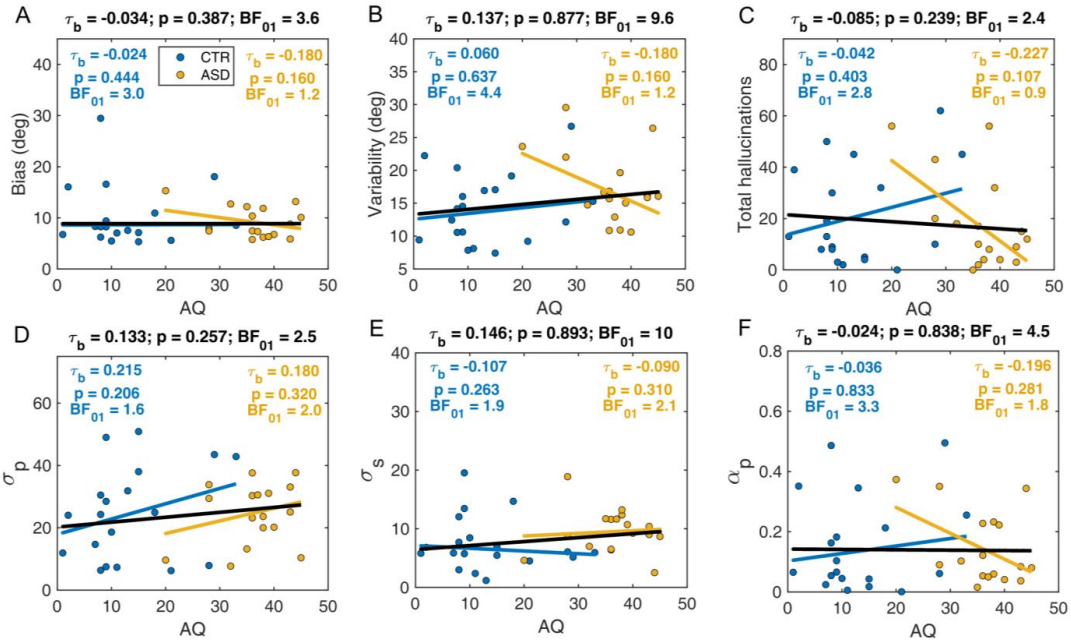
We found that the effects of the acquired priors became significant within 110 trials for all measures, while there were no group differences in any measure at any point in the task (**Supplementary Fig. 3**).



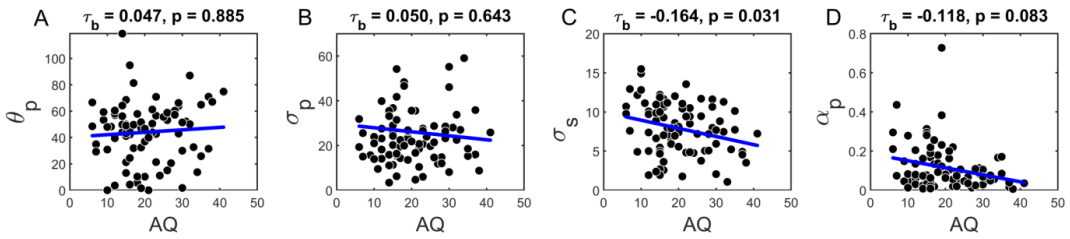
Supplementary Figure 3: Emergence of prior effects in controls (blue boxplots) and ASD participants (yellow boxplots). Cumulative moving averages of median differences between **(A)** estimation RTs at ±32° and RTs at all other directions, **(B)** fraction of detected stimuli at ±32° and fraction detected at all other directions, **(C)** bias at ±64° with respect to bias at ±32°, and **(D) c**umulative moving averages of the probability ratio of hallucinating predominantly around ±32 on no-stimulus trials. The boxplots indicate 25th and 75th percentiles, the black dash in between indicates the median. The significant effects of prior are indicated above each of the plots (one-tailed Wilcoxon signed rank test), while significant group differences are indicated within the plots (two-tailed Wilcoxon rank-sum test). *, ** and *** denote significance levels at $p<0.05$, $p<0.01$, $p<0.001$ respectively.



Supplementary Figure 4: Exploratory analysis of ADOS vs the main measures of interest: (**A, D)** estimation bias, (**B, E**) estimation variability, (**C, F**) the number of hallucinations. (**A-C**) Comparing ADOS negative (orange) and ADOS positive (yellow) subgroups with controls (blue). Statistics reported on each panel are Bayesian Factors for the null hypothesis (BF_{01}) based on one-sided U test: testing the hypothesis that controls show higher values for a given measure. (**D-F**) Correlational analysis using total ADOS score. p-values and BF_{01} denote one-sided Kendal's Tau-b correlation: testing the hypothesis of a negative correlation between ADOS and behavioural measures. Together, the results provide substantial evidence against the hypothesis of a negative relationship between ADOS and the estimation measures of bias and variability. The evidence for a relationship between ADOS and the number of hallucinations, on the other hand, remains inconclusive.



Supplementary Figure 5: Correlational analysis of AQ and the main variables of interest. Behavioural measures: (**A**) estimation bias, (**B**) estimation variability, (**C**) the number of hallucinations. BAYES_P model parameters of interest: (**D**) prior uncertainty, (**E**) sensory uncertainty, (**F**) prior-based lapse rate. Kendal's Tau-b correlation results are reported on each panel for controls (blue) and ASD group (yellow) separately as well as for pooled data (black). For bias, variability, hallucinations and sensory uncertainty the tests were one-tailed while for prior uncertainty and prior-based lapse rate two-tailed test was used.



Supplementary Figure 6: Re-analysis of Karvelis et al. (2018) results using BAYES_P model. (**A**) prior mean, (**B**) prior uncertainty, (**C**) sensory uncertainty, (**D**) prior-based lapse rate. Kendal's Tau-b correlation results and two-sided p-values are reported above each plot. The correlation between AQ and sensory uncertainty is significant (τ_b = -0.164, p = 0.031). This replicates the main finding of Karvelis et al. (2018) that was based on analysing model 'BAYES'. The latter model assumed uniformly distributed lapse estimations, while in BAYES_P lapse estimations followed the acquired prior distribution.

Modelling

Response strategy models ('ADD')

We controlled for the possibility that the task behaviour might be explained by simple behavioural strategies that do not involve Bayesian integration (Laquitaine & Gardner, 2018). This class of models assumed that participants did not combine their expectations with sensory information but relied on either of them alone on any given trial.

The first model, 'ADD1r', assumed that estimations derived from prior expectations were simply sampled from a learnt prior distribution, $p_{prior}(\theta)$, which was parameterized as in Eq (4) - a symmetrical bimodal distribution with nodes at θ_p and $-\theta_p$ and widths of σ_p. However, on trials when participants did perceive motion direction, it was based solely on the sensory input, $p_{likelihood}(\theta_s|\theta_{act}) = V(\theta_{act}, \sigma_s)$.

Putting together the estimations derived from sensory input and the ones derived from learnt expectations, and the possibility of random estimations, the average distribution of estimation responses for a single participant is:

$$p(\theta_{est}|\theta_{act}) = (1-\alpha)\cdot[(1-a)\cdot p_{likelihood}(\theta_s|\theta_{act}) + a\cdot p_{prior}(\theta)] * V(0,\sigma_m) + \alpha, \quad (9)$$

where the asterisk ($*$) denotes convolution and 'a' is the probability that on any given trial the sample will be drawn from the prior; following the winning 'Switching Observer Model' model in Laquitaine & Gardner (2018) 'a' was defined based on the relative precision of the prior: $a = 1/\sigma_p^2 / (1/\sigma_p^2 + 1/\sigma_s^2)$. Just like BAYES and BAYES_P, the resulting 'ADD1r' model had 4 free parameters (θ_p, σ_p, σ_s and α).

The second model, 'ADD2r', was the same as 'ADD1r' except that it had more complex strategy for trials when participants relied on the prior: instead of sampling from the complete acquired prior distribution ranging from −180◦ to +180◦ (Eq. (4)), they sampled only from the negative (−180° to 0°) or the positive (0° to +180°) half, depending on which side of the distribution the actual stimulus occurred on:

$$p_{priorN}(\theta) = V(-\theta_p, \sigma_p) \quad (10)$$

$$p_{priorP}(\theta) = V(\theta_p, \sigma_p) \quad (11)$$

Incorporating this into the distribution of estimation responses results in:

$$p(\theta_{est}|\theta_{act}) = (1-\alpha) \cdot [(1-a) \cdot p_{likelihood}(\theta_s|\theta_{act}) + a \cdot ((1-b(\theta))\, p_{priorN}(\theta) + b(\theta) \cdot p_{priorP}(\theta)] * V(0,\sigma_m) + \alpha, \quad (12)$$

where asterisk (∗) denotes convolution; b (θ) determines the proportion of trials in which participants sample from either negative or positive parts of the prior distribution, respectively; 'b' could take different values for each of the 5 angles: 0◦, ±16◦, ±32◦, ±48◦, ±64◦). The resulting model had 9 parameters.

Finally, we also considered two variations of the 'ADD1r' and 'ADD2r' models. These were identical to 'ADD1r' and 'ADD2r' except from setting σ_p to zero (i.e. no uncertainty in expectations); that is, on trials when perceptual estimates were derived only from expectations, they were equal to the mode of the learnt distribution. This also meant that 'a' was now estimated as a free parameter. These models are referred to as 'ADD1r_m' and 'ADD2r_m'.

Note that in our previous publications (Chalk et al., 2010; Karvelis et al., 2018, Valton et al., 2019) 'a' was an angle-dependent free parameter which could take different values for each of the 5 angles: 0◦, ±16◦, ±32◦, ±48◦, ±64◦), effectively adding 5 extra parameters to each of the ADD* models. The reduced versions of these models (hence the letter 'r') presented in this paper were found to outperform the previous versions of the models and for conciseness the old versions were excluded from model comparison.

Parameter estimation

We used the performance in trials with the highest contrast level to estimate motor noise, σ_m, for each individual. We assumed that at this level sensory uncertainty was close to zero ($\sigma_s \approx 0$). To account for lapse estimations, the motor noise was determined by fitting estimation responses

at the highest contrast level to the distribution in Eq. (2) using the actual motion direction, θ_{act}, as the mean. The estimated motor noise of each individual was used in all the subsequent model fitting as a fixed parameter.

The free parameters of each model were estimated by fitting the response data from the two staircased contrast levels (~200 trials per participant). For each model with a set of free parameters M, we computed the probability distribution $p(\theta_{est} \mid \theta_{act}; M)$ of making an estimate θ_{est} given the actual stimulus direction θ_{act}. For the response strategy models, by definition, the $p(\theta_{est} \mid \theta_{act}; M)$ corresponds to average behaviour in the task (Equations 9 and 12). Bayesian models, on the other hand, explicitly model trial-to-trial variability in the posterior estimate, which in our case is the mean of the posterior (Eq. (6)). To relate this to the behavioural data we built a distribution of 1,000 samples for each presented angle (where each sample is the mean of the posterior obtained via Eq. (6) and perturbed by motor noise via Eq. (7) or (8)).

The parameters were estimated by maximizing the fit of the log likelihood function for the experimental data for each participant individually:

$$M = \underset{M}{\operatorname{argmax}}\left[\sum_i^n \log\left(p\left(\theta_{est} = \theta_{i,data} \mid \theta_i\right)\right)\right], \quad (13)$$

where $\theta_{i,data}$ is participant's estimation response, θ_i is the actual presented motion direction on the ith trial and n is the number of trials. The maximum likelihood was found using *fminsearchbnd* function in Matlab, by minimizing negative log-likelihood. Parameters α, a and b were bounded between 0 and 1, while θ_p, σ_p and σ_s were bounded from 0 to ∞. To reduce the possibility of convergence at local maxima we performed 20 different initializations with parameter values randomly sampled from the range that we found in our previous work (Chalk et al., 2010; Karvelis et al., 2018; Valton et al., 2019).

Model Comparison

To compare the model fits we used Bayesian Information Criterion (BIC), which approximates the log of model evidence (e.g., see Burnham and Anderson, 2004):

$$-2 \cdot \log(P(D|M)) \approx BIC = -2 \cdot \log(P(D|M, \Theta)) + k \cdot \log(n) \quad , \qquad (14)$$

where M is model, D is observed data and P (D|M, Θ) is the likelihood of generating the experimental data given the most likely set of parameters, Θ; k is the number of model parameters and n is the number of data points (or equivalently, the number of trials). BIC evaluates the model by balancing the goodness of fit with model complexity (i.e. the number of model parameters) to avoid over-fitting. Lower BIC score indicates a better model. We also performed a random effect Bayesian model selection analysis (Rigoux et al., 2014). For this, we used the VBA Matlab toolbox (Daunizeau et al., 2014), and used participant-level BIC as an approximation of the log-model evidence as an input.

Parameter recovery

To test the reliability of the parameter estimates of our winning BAYES_P model we performed parameter recovery. The details of this analysis have already been reported in Valton et al. (2019). We found that the winning BAYES_P model recovered parameters very well, which was reflected in the coefficient of determination (R^2) for all recovered parameters being $R^2 \geq 0.84$: prior mean, $R^2 = 0.96$, prior precision, $R^2 = 0.89$, sensory precision, $R^2 = 0.84$, prior-based lapse rate, $R^2 = 0.94$.

References

Burnham, K.P. and Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociological methods & research, 33(2), 261-304.

Chalk, M., Seitz, A.R., Seriès, P., 2010. Rapidly learned stimulus expectations alter perception of motion. Journal of Vision 10, 2.

Daunizeau, J., Adam, V. and Rigoux, L., 2014. VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Computational Biology*, *10*(1), p.e1003441

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191

Rigoux, L., Stephan, K.E., Friston, K.J. and Daunizeau, J., 2014. Bayesian model selection for group studies—revisited. *Neuroimage*, *84*, pp.971-985.

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: a systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular autism*, *6*(1), 2.

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*, e34115.

Laquitaine, S., & Gardner, J. L. (2018). A Switching Observer for Human Perceptual Estimation. *Neuron*, *97*(2), 462–474.

Part II

Additional models of the Moving Dots task

5

Repulsive biases in the Moving Dots task

5.1 Introduction

In all of the analyses presented in the first part of the thesis, as well as in all other previous studies using the same task (Chalk et al., 2010; Gekas et al., 2013), participants exhibited non-zero biases at the most frequent directions. As a result, the modes of the recovered priors also showed a systematic shift away from the most frequent directions (e.g., see **Fig. 6b** and **6d** in Karvelis et al., 2018), suggesting that the recovered priors were capturing additional effects to those of induced expectations for the most frequent directions. This exemplifies one of the main dangers of Bayesian models: any non-Bayesian task effects will inevitably be explained by the prior if not explicitly accounted for (e.g., see Bowers and Davis, 2012). The goal of this chapter is to account for these effects and reassess the individual and group differences in our data.

We speculated that these additional effects might originate from other pre-existing priors, possibly relating to the global coordinates. In our task, the stimuli

motion directions are set relative to the central reference angle, which is randomly chosen for each individual. Randomizing this angle is meant to ensure that any potential effects arising from the global coordinates - i.e., with respect to the cardinals (horizontal and vertical directions) - are averaged out when analysing group data. However, this might not completely work. The two most frequent directions divide the 360° circle into two unequal regions: one of 64° and another of 296°. This means that despite selecting the central reference angle randomly, the cardinal directions are more likely to fall within the larger region thus exerting more effects in this region when averaging across individuals.

This, however, still does not explain how the cardinal directions are affecting performance in our task. One possibility could be that people have a prior for cardinal directions. One of the classical examples of perceptual priors is the prior for cardinal orientations (Tomassini et al., 2010; Girshick et al., 2011). These studies showed that the biases that people exhibit when making perceptual judgements about Gabor stimuli orientation can be well explained by a prior for cardinal directions. Furthermore, they showed that such prior corresponds to natural scene statistics. While these reports are constrained to orientation judgements only, we might expect to find similar priors in motion direction judgements.

Another possibility might relate to what is known as *reference repulsion*. Rauber and Treue (1998) studied motion direction judgements using random-dot patterns (RDPs) and found repulsive biases (of up to 9°) from the cardinal directions - which were not explicitly marked in any way. Similarly, in a different condition the authors used an explicit cue - a line segment oriented to a particular direction - presented together with the stimuli and found repulsive biases from this cue too. This suggested an interpretation that cardinal directions are used as implicit reference directions.

While the exact mechanism via which reference repulsion arises is still not completely understood, a few accounts have been put forward. Jazayeri and Movshon (2007) investigated reference repulsion using an explicit cue and proposed a non-Bayesian model where the repulsion originates at the decoding stage via a non-uniform read-out profile of the sensory likelihood, which gives higher weight to more informative neuron populations. Wei and Stocker (2015) presented a Bayesian model in which the repulsion originates from the encoding stage via efficient coding of the sensory likelihood. Contrary to Wei and Stocker model, however, Zamboni et al. (2016) demonstrated reference repulsion to be a post-perceptual bias. Zamboni and colleagues examined reference repulsion in two conditions: 1) when the the reference cue was present during both stimulus presentation and estimation stage and 2) when the cue was only present during stimulus presentation but not during the estimation. The repulsive biases were observed only in the first condition suggesting that reference repulsion happens at the decision-making stage and is not perceptual. However, Luu and Stocker (2018) re-analyzed this data with their proposed model and found that reference repulsion, while much weaker, was still present in the second condition. In Luu and Stocker model the repulsion arises from a post-decision *self-consistency principle* applied to perceptual inference. According to this principle, when stimulus estimation is preceded by a discrimination/categorical judgement (either explicit or implicit), it truncates the part of the posterior distribution that extends outside of the selected category (i.e. setting the posterior probability to zero in that region). Such truncation effectively shifts the mean of the posterior, giving rise to repulsive biases. Fritsche and de Lange (2019) performed a series of experiments that further elaborated on such mechanisms. Their results suggested that the preceding discrimination judgement leads to repulsion effects via working memory representations, thus further supporting the idea of reference repulsion being

of non-perceptual nature (although see Kuang (2019) for an alternative view).

In what is presented below, we incorporate the above ideas and models in trying to explain the unaccounted biases observed in our data. Note, however, that most of these studies investigated reference repulsion by using explicit reference cues, while in our task only implicit reference directions are possible. This adds an additional element of uncertainty as to what mechanism we might expect to underlie the effects observed in our task. For that reason we aimed to be comprehensive and did not rule out any of these proposals a priori. Apart from finding a model that fits the data well we placed a particular focus on: 1) examining whether the results of this new model changes the interpretation of our previous analyses and 2) assessing individual and group differences in relation to the additional effects captured by the model.

5.2 Methods

5.2.1 Behavioral data analysis

Before any modelling, we re-examined the behavioral data to gain insight into what these additional effects might be. We remapped the behavioral data from the relative coordinates as it had been in all of the analysis thus far - i.e. coordinates relative to the central reference angle - into the absolute coordinates. To achieve this, we binned the response data by dividing the 360° circle into 45 bins of size 8°. This meant that for any given bin only $\sim 36\%$ of participants had any data for it - that is because each participant was tested within a range of 128° around their central reference angle (between $-64°$ and 64°), which is $\sim 36\%$ of the whole range. This also meant that the group averages were ~ 3 times noisier than in our previous analysis. Nevertheless, this averaged out the effects of the acquired priors and left us with the effects arising from the absolute coordinates.

5.2.2 Models

Our modelling approach started with establishing a baseline model. For that, we wanted a model in which the prior reflected only the acquired prior for the two most frequent directions ($\pm 32^\circ$). Since our previous winning model BAYES_P already captured not only the prior for $\pm 32^\circ$ but the combined influence of all the other effects that were at play, it was not suitable for this purpose. Instead, we constructed a reduced version of BAYES_P, where we fixed the mean of the prior to be at $\pm 32^\circ$, because that is where we would expect the mean of the acquired prior to be in the absence of any other effects. We abbreviated this model *BP-A* (i.e. BAYES_P - Acquired prior only). In all other respects the implementation of BP-A was the same as BAYES_P, the full details of which can be found in **Chapter 4**. For clarity and ease of reference, however, a brief description of the model implementation is provided below.

The prior in BP-A was parameterised as two von Mises distributions with modes fixed at $\pm 32^\circ$ and with widths of σ_p:

$$p_{prior}(\theta) = \frac{1}{2}[V(-32, \sigma_p) + V(32, \sigma_p)] \quad , \tag{5.1}$$

As in the previous models, the likelihood was parameterised as a von Mises distribution with width σ_s:

$$p_{likelihood}(\theta_s|\theta) = V(\theta_s, \sigma_s) \quad , \tag{5.2}$$

where θ_s was drawn from another von Mises distribution centred on the actual presented motion direction θ_{act} with the same variance σ_s (i.e. $V(\theta_{act}, \sigma_s)$) to account for trial-to-trial variability.

The posterior of the stimulus motion direction, θ, was then be computed via:

$$p_{posterior}(\theta|\theta_s) \propto p_{prior}(\theta) \cdot p_{likelihood}(\theta_s|\theta) \quad . \tag{5.3}$$

The perceived direction, θ_{perc}, was taken to be the mean of the posterior distribution:

$$\theta_{perc} = \int \theta \cdot p_{posterior}(\theta|\theta_s) \; d\theta \tag{5.4}$$

Finally, by accounting for motor noise, σ_m, and prior-based lapse estimations, α_p, the probability distribution of estimation responses, θ_{est}, was obtained via:

$$p(\theta_{est}|\theta_{perc}) = (1 - \alpha_p) \cdot V(\theta_{perc}, \sigma_m) + \alpha_p \quad . \tag{5.5}$$

BP-A had a total of 3 free parameters: σ_p - prior uncertainty, σ_s - sensory uncertainty and α_p - prior-based lapse rate. Next, we considered a variety of models that built on top of BP-A, to account for the additional effects.

5.2.2.1 Efficient coding model: BP-AEC

First, we augmented BP-A by considering efficient coding of the sensory representations as proposed by Wei and Stocker (2015). Efficient coding in this case is assumed to be achieved by maximizing mutual information between the stimulus variable and its sensory representation. Under this constrain, the square root of Fisher information can be shown to be proportional to the prior distribution of the stimulus ($\sqrt{J(\theta)} \propto p(\theta)$). We can then map stimulus space to sensory space by computing the cumulative probability of the prior distribution:

$$F(\theta) = \int_{-\infty}^{\theta} p(\chi) d\chi \; , \tag{5.6}$$

where we use a uniform prior and use notation $\tilde{\theta} = F(\theta)$ to denote the sensory space variable. We then perform inverse mapping, $\theta = F^{-1}(\tilde{\theta})$, back to the stimulus space using the prior distribution that a participant had in our task. This results in sensory representations that are asymmetric, with a longer tail away from the most expected directions (**Fig. 5.1**) - which sets the stage for repulsive biases.

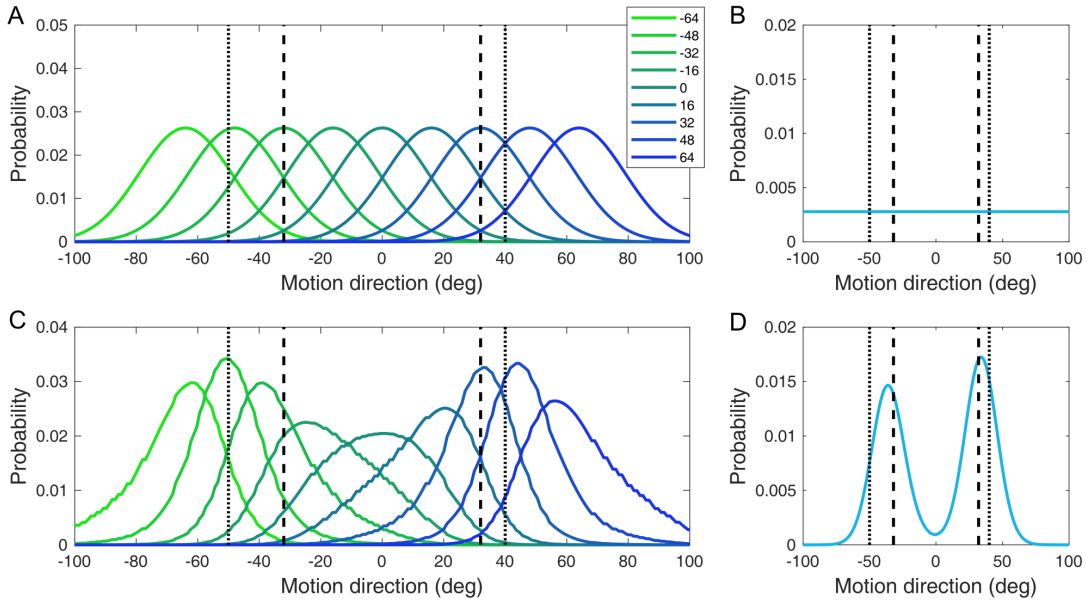


Figure 5.1: An example of the transformation of sensory representations following efficient coding. (**A**) Sensory likelihoods for different presented motion directions in the sensory space or under the assumption of a uniform prior distribution (**B**). (**C**) Sensory likelihoods for different motion directions in the stimulus space after transformation using the prior in our task (**D**) - which was assumed to be the combined prior of the most frequent and cardinal directions. The dashed lines indicate the two most frequent directions, while the dotted lines indicate the cardinal directions.

The transformations were performed according to a combined prior which was a product of the acquired prior and the cardinal prior. The assumption of the cardinal prior for this model was necessary so as to make it possible to account for the repulsive biases with respect to the cardinals seen in the behavioral data. The cardinal prior was parameterised as 4 von Mises distributions centered on -90, 0, 90 and 180:

$$p_{cardinal}(\theta) = \frac{1}{4}[V(-90 - z_r,\ \sigma_c) + V(0 - z_r,\ \sigma_c) + V(90 - z_r,\ \sigma_c) + V(180 - z_r, \sigma_c)] \ , \tag{5.7}$$

where the width around each peak is a free parameter σ_c, and z_r is the central reference angle (unique to each participant), which here serves to convert the cardinal prior from absolute to relative coordinates. The combined prior was simply:

$$p_{combined}(\theta) = p_{prior}(\theta)\, p_{cardinal}(\theta) \ , \tag{5.8}$$

The posterior of motion direction on a single trial then can be expressed as:

$$p_{posterior}(\theta|\theta_s) \propto p_{likelihood}(\theta_s|F^{-1}(\tilde{\theta}))\, p_{combined}(F^{-1}(\tilde{\theta})) \ , \tag{5.9}$$

Apart from this transformation, and replacing Eq. 5.3 with Eq. 5.9, the implementation of this model was exactly the same as BP-A and it had one additional free parameter - the width of the cardinal prior, σ_c. We termed this model *BP-AEC*.

5.2.2.2 Self-consistency models: BP-ASC4 and BP-ASC40

Another model we tested was based on the self-consistency principle proposed by Luu and Stocker (2018). We reasoned that in our task, before making a precise estimate of the stimulus motion direction direction, participants might be first making an implicit categorical judgement, $C \in \{c_1, c_2, c_3, c_4\}$, about the quadrant of the motion direction - the boundaries of these quadrants being at the cardinal directions. This preceding judgement would truncate the part of the posterior distribution that falls outside of the judged quadrant. The truncation would result in the mean of the posterior shifting away from the boundary of that quadrant,

giving rise to repulsive biases (**Fig. 5.2**).

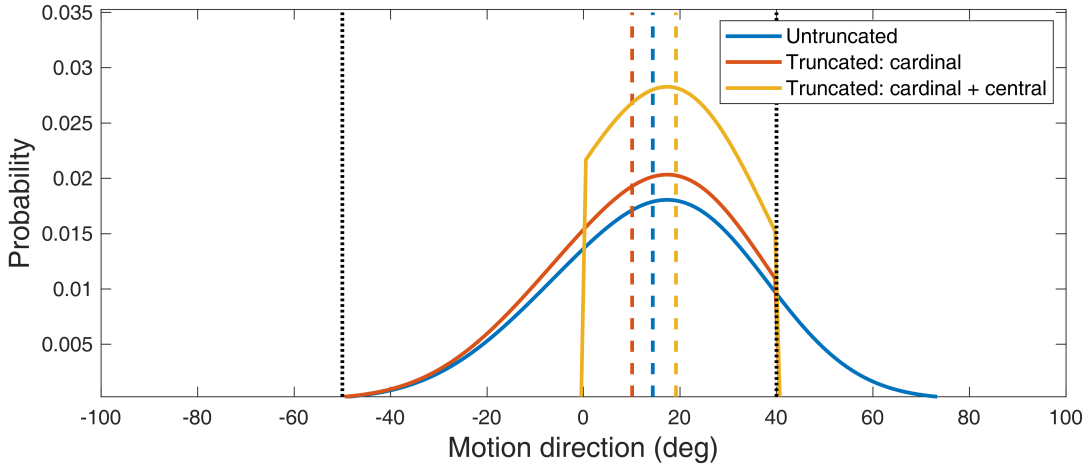


Figure 5.2: An illustration of how truncation of the posterior affects its mean estimate. Blue line - the untruncated posterior. Red - posterior truncated at the boundaries of the judged quadrant. Yellow line - posterior truncated at the boundaries of the quadrant and at the central reference angle. The colored dashed vertical lines indicate the means of the corresponding distributions. The dotted lines indicate cardinal directions.

In computational terms, a categorical judgement, $\hat{C}$, has to be made based on a noisy sensory estimate θ_s of the actual motion direction θ_{act}. For a symmetric loss function, this means choosing the category with the highest posterior probability; as we assumed the prior $p(C)$ to be uniform, it was equivalent to choosing the category with the highest likelihood:

$$\hat{C} = \arg\max_C \; p(C|\theta_s) \;=\; \arg\max_C \; p(\theta_s|C) \;. \tag{5.10}$$

The likelihood was obtained by marginalizing over all motion directions:

$$p(\theta_s|C) = \int p(\theta_s|\theta)\, p(\theta|C) d\theta \;, \tag{5.11}$$

where $p(\theta_s|\theta)$ was defined as in the previous models as a von Mises distribution with a mean θ and variance σ_s. $p(\theta|C)$ was defined as a boxcar function which is uniform within an interval that spans a given category and is zero outside of

this interval.

The posterior of motion direction on a given trial is then conditioned on the categorical judgement:

$$p(\theta|\theta_s, \hat{C}) \propto p(\theta|\theta_s) \cdot \Pi_{\hat{C}}(\theta) \quad , \tag{5.12}$$

where $\Pi_{\hat{C}}$ is a boxcar function with boundaries determined by the estimated category $\hat{C}$.

Apart from the categorical dependence described here, and replacing Eq. 5.3 with Eq. 5.12, the model was implemented in the same way as BP-A and it had the same 3 free parameters. We termed this model *BP-ASC4* for a self-consistency principle stemming from the 4 cardinal directions.

Since our initial analysis showed that the effects of truncation at cardinals were insufficient to account for the additional biases around the central reference angle, we hypothesised that participants might be making an additional categorical judgement, $C' \in \{c_-, c_+\}$, for the stimulus being counterclockwise, c_-, or clockwise, c_+, from the central reference angle, leading to further truncation of the posterior (**Fig. 5.2**). In mathematical notation this can be expressed as:

$$p(\theta|\theta_s, \hat{C}, \hat{C}') \propto p(\theta|\theta_s) \cdot \Pi_{\hat{C}}(\theta) \cdot \Pi_{\hat{C}'}(\theta) \quad , \tag{5.13}$$

where we assumed the estimate of the category, $\hat{C}'$, to be made in the same way as described in Eq. 5.10. Thus, the final model, *BP-ASC40*, truncated the posterior both at the boundaries of the quadrant and at the central reference angle. BP-ASC40 had the same 3 free parameters as BP-A.

5.2.2.3 Non-mechanistic repulsion models: BP-AR4 and BP-AR40

Finally, as the models considered above did not account for the repulsive effects well (see **Section A.2 in Appendix**), we resorted to implementing models that are agnostic to the mechanisms producing repulsive biases, but instead model them explicitly. Such models, again, were equivalent to BP-A, but had a repulsive bias added at the post-perceptual stage. First, we considered repulsive biases from the cardinal directions. This was implemented as a periodic linear function of motion direction (see **Fig. 5.3**), which in absolute coordinates can be expressed as:

$$b_4(\theta) = \begin{cases} (1 - \frac{4}{\pi}(\theta + \pi))\, r_4 & if -180 < \theta < -90 \\ (1 - \frac{4}{\pi}(\theta + \frac{\pi}{2}))\, r_4 & if -90 < \theta < 0 \\ (1 - \frac{4}{\pi}\theta)\, r_4 & if\ 0 < \theta < 90 \\ (1 - \frac{4}{\pi}(\theta - \frac{\pi}{2}))\, r_4 & if\ 90 < \theta < 180 \\ 0 & if\ \theta \in \{180,\ 90,\ 0,\ -90\} \end{cases} \tag{5.14}$$

where r_4 is a free parameter that captures the amplitude of repulsive bias from the cardinals. These effects entered the model at Eq. 5.5, resulting in:

$$p(\theta_{est}|\theta_{perc}) = (1 - \alpha_p) \cdot V(\theta_{perc} + b_4(\theta_{act} + z_r), \sigma_m) + \alpha_p \quad , \tag{5.15}$$

where the central reference angle, z_r, again serves to convert the effects from absolute to relative coordinates. In all other aspects this model was implemented as BP-A. It had one additional free parameter, r_4. We termed this model *BP-AR4*.

Similarly to BP-ASC40, we also considered a model where in addition to repulsion from the cardinals, there was a repulsion from the central reference

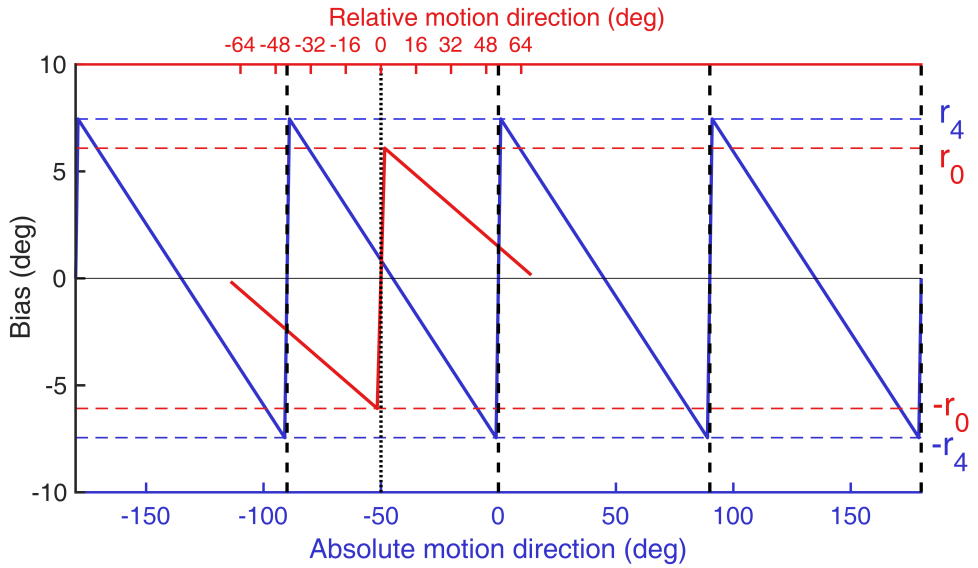


Figure 5.3: An illustration of post-perceptual repulsive biases functions from the cardinals (b_4; blue line; Eq. 5.14) and from the central reference angle (b_0, red line; Eq. 5.16) that were used for models BP-AR4 (only b_4) and BP-AR40 (both b_4 and b_0). r_4 and r_0 control the amplitude of each function (b_4 and b_0, respectively) and were included in the models as free parameters.

angle. This was also implemented as a linear function of motion direction. For simplicity, we assumed the slope of this function to be such that the bias would go to 0 at the boundary of the tested motion range ($\pm 64^\circ$) (see **Fig. 5.3**):

$$b_0(\theta) = \begin{cases} -(1+\theta\frac{16\pi}{45})\,r_0 & if\ \theta < 0 \\ (1-\theta\frac{16\pi}{45})\,r_0 & if\ \theta > 0 \\ 0 & if\ \theta = 0 \end{cases} \tag{5.16}$$

with an additional parameter, r_0, to capture the amplitude of the bias from the central reference angle. The distribution of estimated motion directions was then obtained by replacing Eq. 5.5 with:

$$p(\theta_{est}|\theta_{perc}) = (1-\alpha_p)\cdot V(\theta_{perc}+b_4(\theta_{act}+z_r)+b_0(\theta_{act}),\sigma_m)+\alpha_p \quad , \tag{5.17}$$

This model was termed *BP-AR40*, and it had a total of 5 free parameters. All models are summarized in **Table 5.1**.

Table 5.1: Summary of all repulsion models and their relationships. Parameter meanings: σ_p - prior uncertainty, σ_s - sensory uncertainty, α_p - prior-based lapses, σ_c - cardinal prior uncertainty, r_4 - amplitude of repulsion from the cardinals, r_0 - amplitude of repulsion from the central reference angle

Model	Descrption	Free parameters
BP-A	A baseline model with the acquired prior only: identical to BAYES_P but with prior mean fixed at 32°	σ_p, σ_s, α_p
BP-AEC	Like BP-A, but with an added cardinal prior and efficient coding of sensory likelihood (Wei and Stocker, 2015)	σ_p, σ_s, α_p, σ_c
BP-ASC4	Like BP-A, but with truncation of the posterior distribution at cardinal directions following the self-consistency principle (Luu and Stocker, 2018)	σ_p, σ_s, α_p
BP-ASC40	Like BP-ASC4, but with an additional truncation of the posterior distribution at the central reference angle	σ_p, σ_s, α_p
BP-AR4	Like BP-A, but with a repulsive bias from the cardinals added at the post-perceptual stage	σ_p, σ_s, α_p, r_4
BP-AR40	Like BP-AR4, but with an additional repulsive bias from the central reference angle added at the post-perceptual stage	σ_p, σ_s, α_p, r_4, r_0

5.2.3 Model fitting

Model fitting followed the same procedure as for the Bayesian models in the previous chapters. The models were fit by maximizing the log likelihood function for the motion estimation data for each participant individually. The maximum likelihood was found using *fminsearchbnd* function in Matlab, by minimizing negative log-likelihood. α_p was bounded between 0 and 1, while σ_p, σ_s, r_4 and r_0 were bounded from 0 to ∞. To reduce the possibility of convergence at local maxima we performed 20 different initializations with parameter values sampled randomly within a range found in previous studies (Karvelis et al., 2018; Valton et al., 2019) for parameters α_p, σ_p and σ_s, while for r_4 and r_0 the range was [0 25] and [0 10], respectively. The latter were based on the maximum range determined after an initial unconstrained exploration of the parameter space.

5.2.4 Model comparison

As in the previous chapters, to compute the log of model evidence we used Bayesian Information Criterion (BIC) approximation:

$$\log(P(D|M)) \approx -\frac{BIC}{2} = \log(P(D|M,\hat{\Theta})) + k \cdot \log(n) \tag{5.18}$$

where M is model, D is observed data and P (D|M,$\hat{\Theta}$) is the likelihood of generating the experimental data given a set of parameters,$\hat{\Theta}$; k is the number of model parameters and n is the number of data points (i.e. trials). To compare the models we performed a random-effect Bayesian model selection (BMS) analysis (Stephan et al., 2009b) using VBA Matlab toolbox (Daunizeau et al., 2014). In random-effect BMS, model comparison is based on exceedance probability (EP), which expresses the likelihood that any given model is more frequent than all other models. Correcting for chance variations in model frequencies (Rigoux et al., 2014), we can obtain protected exceedance probability (PEP):

$$PEP_k = EP_k(1 - BOR) + \frac{1}{K}BOR \quad , \tag{5.19}$$

where subscript k denotes kth model from the total of K models being compared; BOR is Bayesian Omnibus Risk, which quantifies the chance likelihood of the observed model frequencies:

$$BOR = \frac{p(y|H_0)}{p(y|H_0) + p(y|H_1)} \quad , \tag{5.20}$$

where y contains log model evidences (as computed in **Eq. 5.18**) of all subjects and all models, H_0 assumes that model frequencies are all the same and H_1 assumes that at least one of the models is more frequent. Thus, the higher the BOR estimate, the less differences there will be in PEP of different models.

Since random-effects BMS was often inconclusive, we also performed fixed-effects model comparison. For this we summed the BIC across all participants and used this sum to compare the models. Unlike random-effects model comparison, this assumes that all participants in the group are best fit by the same model.

5.2.5 Model and parameter recovery

Modelling results, and their robustness, can be affected by many factors, such as parameterisation of models, amount of data used for model fitting, chosen model fitting methods, etc. A general way to account for these factors and to assess the overall reliability of modelling results is to perform model and parameter recovery (e.g., see Palminteri et al., 2017). Given the constraints of the mentioned factors, model recovery shows the distinguishability of models, while parameter recovery shows reliability of parameter estimates.

For each model, we generated 100 sets of parameters (100 synthetic individuals), by sampling the distributions of parameter values found in our general population sample. We then simulated 200 responses for each synthetic participant (just as we had ∼200 responses per participant in our data). Every model was then fitted to every simulated dataset following the same model fitting procedure as described in the previous section. We then performed model comparison for every simulated dataset, also as described in the previous section.

We also analysed how well the parameters were recovered by the winning model. This was done by comparing parameters that simulated the data ('actual parameters') with the parameters estimated from the simulated data ('recovered parameters') using the coefficient of determination (R^2) obtained from *regress* function in Matlab.

5.2.6 Datasets

The behavioral data analysis of repulsion effects and most of the modelling analysis was done using the general population data ($N = 83$) from **Chapter 2**, as this was our largest sample. Then we used the same modelling approach to investigate SCZ sample (from **Chapter 3**) and ASD sample (from **Chapter 4**).

5.2.7 Statistical tests

The behavioral performance effects with respect to the absolute coordinates were tested using a linear mixed model in SPSS version 25. Whenever there was periodicity in the effects, the data was folded accordingly to reduce the noise before running the test: e.g., folding the data with respect to 0° would mean averaging data points at an equal absolute distance from the 0° angle. Due to the presence of outliers and non-normalities in the data, correlation analysis of personality traits and model parameters was done using Kendall's Tau-b correlation. Wherever relevant, the evidence for the null hypothesis was evaluated by computing Bayes factors (BF_{01}) using JASP version 0.10 (JASP Team, 2019), with the default Cauchy prior of scale 0.707. Bayes factor below 3 is generally considered inconclusive, above 3 is considered moderate evidence, above 10 is strong evidence and above 30 is very strong evidence (Kass and Raftery, 1995).

5.3 Results

5.3.1 Behavioral results

Analyzing estimation biases with respect to the absolute coordinates revealed strong biases (up to $\sim 10^{\circ}$) that exhibited periodicity tied to the cardinal directions (**Fig. 5.4**). These effects were highly significant (the main effect of absolute motion direction: $F(11, 83.04) = 10.10,\ p < 10^{-5}$) and suggested repulsive biases from the cardinal directions.

5.3.2 Modelling results

We fitted and compared five models that built on top of our baseline model BP-A. We computed log model evidence using BIC approximation and used

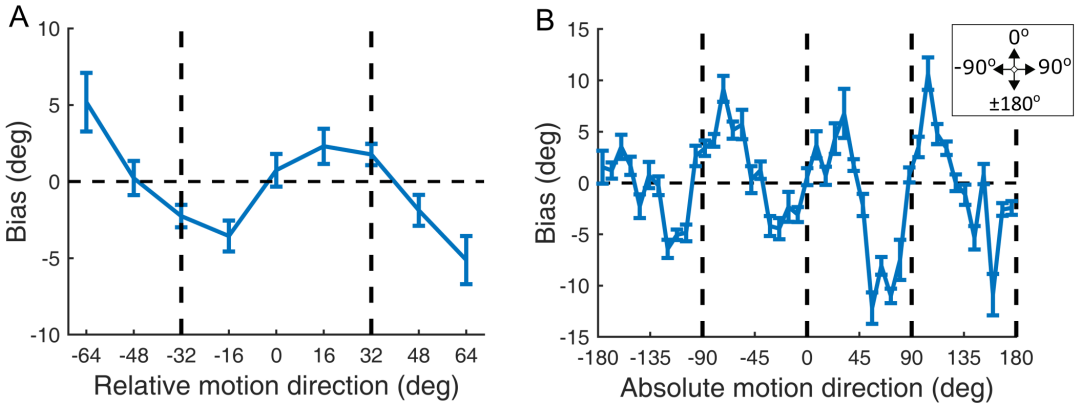


Figure 5.4: Average group ($N = 83$) biases on low contrast trials. (**A**) Biases calculated with respect to the central reference angle that was randomly selected for each individual. The vertical dashed lines indicate the two most frequently presented motion directions. (**B**) Biases calculated with respect to the absolute coordinates. The vertical dashed lines indicate the cardinal directions. The error bars represent within-subject standard error. For other behavioral measures with respect to absolute motion direction coordinates see **Section A.1 in Appendix**.

both random-effects BMS analysis and fixed-effect analysis to compare the models. Random-effects BMS was not completely conclusive, with the best model, BP-AR40, having PEP = 0.21 (**Fig. 5.5A**); before BOR correction, however, BP-AR40 was chosen with EP = 0.74. Furthermore, fixed-effects analysis also favoured BP-AR40 model, it being better than the next best model (BP-AR4) by 96 BIC units (**Fig. 5.5B**).

Comparing the fits of the two best-fitting models BP-AR4 and BP-AR40, and the baseline model BP-A, we can observe that while repulsion from the cardinals was fully accounted for by BP-AR4 (orange line, **Fig. 5.5D**), these effects only partially accounted for the repulsive biases observed relative to the central reference angle (orange line, **Fig. 5.5B**). An additional repulsion from the central reference angle - model BP-AR40 - provided a much better fit to the biases in these coordinates (purple line, **Fig. 5.5C**) while accounting for repulsion from the cardinals just as well (purple line, **Fig. 5.5D**).

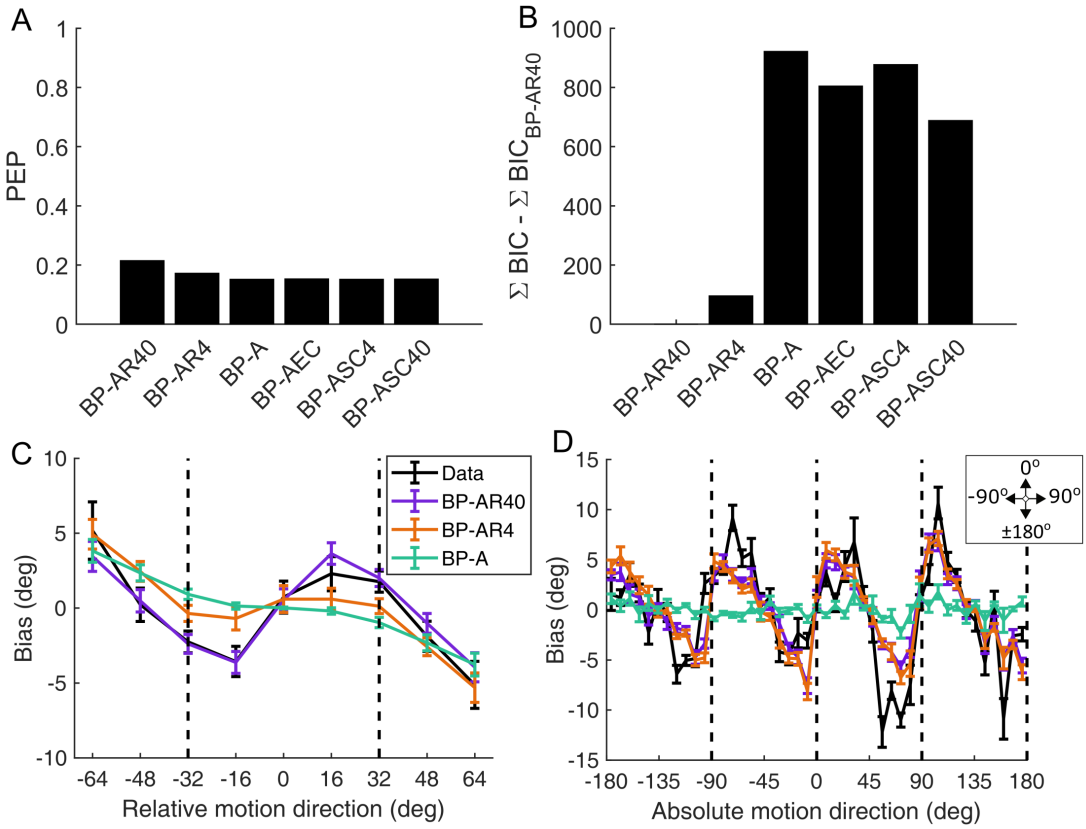


Figure 5.5: Model comparison and model fits for repulsion analysis. (**A**) Random-effects Bayesian model selection results; PEP - protected exceedance probability. (**B**) Fixed-effects model comparison. Note that the plot shows the difference between the model with the lowest BIC and the remaining models. BP-AR40 had the lowest BIC with 96 units lead. (**C**) Model predictions of bias relative to the central reference angle by the two best fitting models (BP-AR40 and BP-AR4) and the baseline model (BP-A). The vertical dashed lines indicate the two most frequently presented motion directions. (**D**) Predictions of the same models for biases relative to the absolute coordinates.The vertical dashed lines indicate the cardinal directions. The error bars represent within-subject standard error. See **Section A.2 in Appendix** for complete model fits of all models.

5.3.2.1 BP-AR40 parameters vs traits

Next, we studied how variation in BP-AR40 model parameters related to autistic (AQ) and schizotypy (SPQ and RISC) traits. AQ scores were found to negatively correlate with sensory uncertainty ($\tau_b = -0.184, p = .015$; **Fig. 5.6A**) - in agreement with our previous modelling analysis using BAYES_P, which ignored the repulsion effects. However, BP-AR40 also suggested a significant negative relationship between AQ and prior-based lapses ($\tau_b = -0.155, p = .041$; **Fig. 5.6B**),

which had not reached significance in BAYES_P model ($p = 0.083$; see **Supplementary Fig. 6** in **Chapter 4**). No other correlations with AQ were significant (prior uncertainty, σ_p : $\tau_b = 0.07$, $BF_{01} = 4.6$; repulsion from cardinals, r_4 : $\tau_b = -0.002$, $BF_{01} = 6.9$; and central reference angle, r_0 : $\tau_b = -0.038$, $BF_{01} = 6.1$).

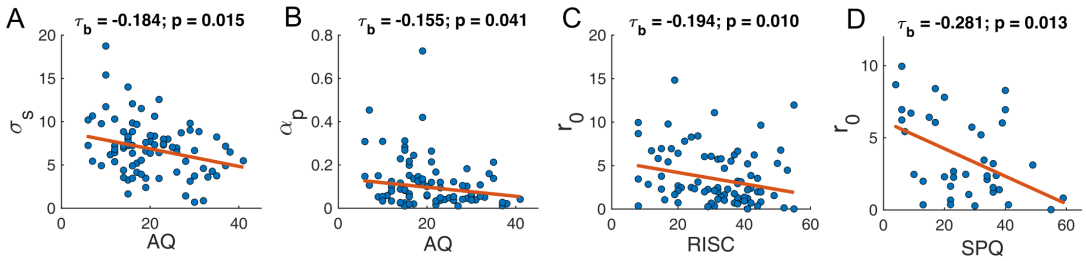


Figure 5.6: Significant results of BP-AR40 model parameter analysis. (**A**) σ_s - uncertainty of the prior distribution vs AQ. (**B**) α_p - prior-based lapse rate vs AQ. (**C**) r_0 - repulsion from the central reference angle vs RISC (**D**) r_0 - repulsion from the central reference angle vs SPQ (N=39). Two-tailed Kendall's Tau-b correlations are reported above each panel.

For schizotypy traits we found a negative correlation with the repulsion from the central reference angle. This correlation was found using both RISC (N=83; $\tau_b = -0.194$, $p = .01$; **Fig. 5.6C**) and SPQ (N=39; $\tau_b = -0.281$, $p = .013$; **Fig. 5.6D**) measures. No other correlations were significant for both RISC (prior uncertainty, σ_p : $\tau_b = 0.001$, $BF_{01} = 6.9$; sensory uncertainty, σ_s : $\tau_b = 0.071$, $BF_{01} = 4.5$; prior-based lapses, α_p : $\tau_b = 0.124$, $BF_{01} = 1.8$, $p = .1$; repulsion from cardinals, r_4 : $\tau_b = 0.047$, $BF_{01} = 5.8$) and SPQ (prior uncertainty, σ_p : $\tau_b = 0.098$, $BF_{01} = 3.3$; sensory uncertainty, σ_s : $\tau_b = -0.035$, $BF_{01} = 4.6$; prior-based lapses, α_p : $\tau_b = -0.046$, $BF_{01} = 4.4$; repulsion from cardinals, r_4 : $\tau_b = 0.145$, $BF_{01} = 2.1$, $p = .2$).

5.3.2.2 Model comparison for SCZ and ASD groups

We also fitted the models to our other datasets: SCZ (described in **Chapter 3**) and ASD (described in **Chapter 4**) to analyze group differences between controls and clinical groups. Random-effect BMS results were mostly inconclusive for the

patient groups with BOR being 0.99 in SCZ and 0.96 in ASD - meaning that differences in model frequencies were likely due to chance, hence, PEP was very similar for all models. For controls, however, BP-AR40 (PEP = 0.40) and BP-AR4 (PEP = 0.33) showed a clear advantage over other models (**Fig. 5.7A-C**).

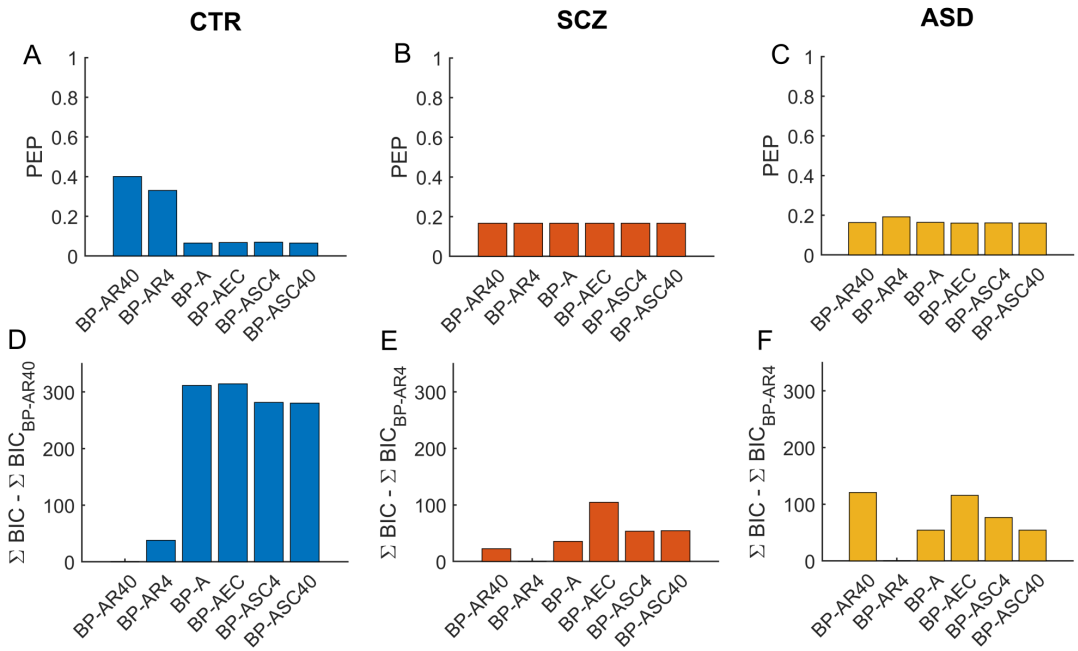


Figure 5.7: Model comparison for patient groups. (**A**-**C**) Random-effects Bayesian model selection; PEP - protected exceedance probability. (**D**-**F**) Fixed-effects model comparison. Note that the plots show the difference between the model with the lowest BIC and the remaining models. (**A**,**D**) Controls, (**B**,**E**) participants with schizophrenia, (**C**,**F**) participants with ASD. Fixed effects analysis showed BP-AR40 to be the winning model in the control group (with 73 BIC units difference from the next best fitting model) while for the patient groups the winning model was BP-AR4, winning by 31 BIC units in SCZ group and 65 BIC units in ASD group.

Due to the inconclusive results with random-effects BMS, we performed fixed-effects model comparison by summing BIC across participants (**Fig. 5.7D-E**). In the control group, BP-AR40 was found to be the winning model, with 38 BIC units difference from the next best model. In SCZ and ASD groups, a simpler model BP-AR4 won the model comparison by 23 and 121 BIC units lead, respectively. Note that anything above 10 BIC units difference is considered very

strong, decisive evidence (Kass and Raftery, 1995). These results suggest that while controls exhibited repulsion from both cardinal directions and the central reference angle, patient groups showed only repulsion from the cardinals. For SCZ group this is in agreement with reduced repulsion from the central reference angle along schizotypy traits in the general population.

5.3.2.3 Model and parameter recovery

Model recovery results showed that most models can be reliably recovered (**Fig. 5.8**). However, the winning model in our analysis, BP-AR40, which assumed repulsion from the cardinals and repulsion from the central reference angle was not recovered perfectly (with PEP = 0.21); the data generated by this model was better explained by a simpler BP-AR4 model, which assumed only repulsion from the cardinals. This suggests, that the assumed repulsion from the central reference angle was not strong enough for most individuals to justify BP-AR40 model. Using fixed-effects model comparison, however, we find BP-AR40 to be recovered well, having 140 BIC units less than BP-AR4.

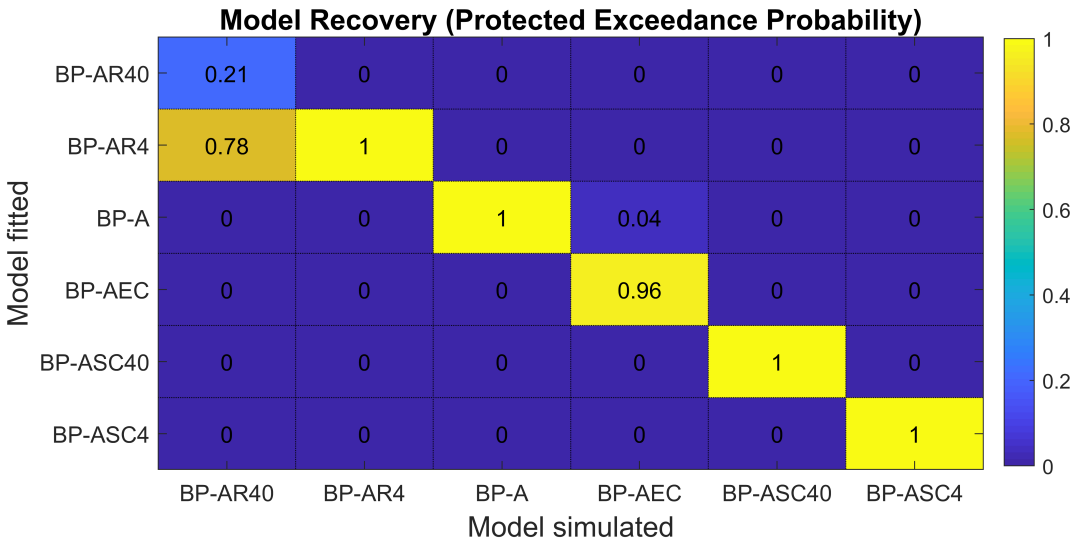


Figure 5.8: Model recovery based on protected exceedance probability. X-axis - models that simulated the data, Y-axis - models that were fitted to the simulated data.

To investigate if the repulsion from the central reference angle would be better captured with more data, we performed the same modelling analysis on an independent dataset (N=12) from Chalk et al. (2010), where each participant had ~700 trials available for model fitting (~3 times more than in our data). While fixed-effects model comparison again favoured BP-AR40, increasing confidence in the result that repulsion from the central angle is significant in the general population, random-effects BMS was still inconclusive (see **Section A.3 in Appendix**).

To assess the reliability of BP-AR40 model results, we performed parameter recovery. All parameters were recovered well, with $R^2 > 0.55$ (**Fig. 5.9**).

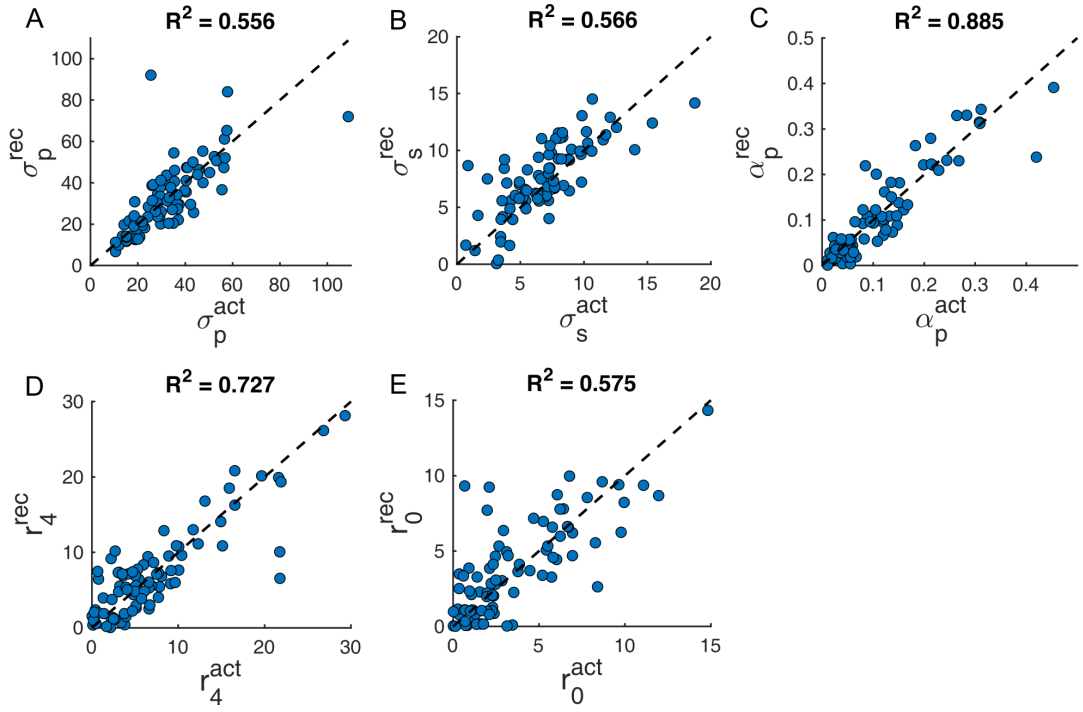


Figure 5.9: Parameter recovery for BP-AR40 model. (**A**) σ_p - uncertainty of the prior distribution, (**B**) σ_s - uncertainty of the sensory likelihood, (**C**) α_p - prior-based lapse rate, (**D**) r_4 - maximum repulsion from the cardinals, (**E**) r_0 - maximum repulsion from the central reference angle. X-axes – actual parameters used for simulating the data (superscript 'act'); Y-axes – recovered parameters (superscript 'rec') after fitting the model to the simulated data. The dashed diagonal line is a reference line indicating perfect parameter recovery. The coefficient of determination is reported above each panel.

5.4 Discussion

In this chapter we sought to investigate the repulsive biases present in our data, which had been ignored in the previous analyses. Reanalysing the motion estimation data with respect to the absolute coordinates showed strong repulsion from the cardinals (up to $\sim 10^{\circ}$). Our aims were: 1) to construct a model that captures these repulsive biases well, 2) to determine if after accounting for these effects the results are consistent to our previous analyses and 3) to assess individual and group differences in relation to the additional effects captured by the model. To this end, we constructed and tested a range of models. BP-AEC model assumed that repulsive biases arise from efficient coding of sensory likelihood (Wei and Stocker, 2015), while BP-ASC models assumed repulsive biases to arise from truncation of the stimulus direction posterior distribution due to a preceding implicit categorization of the stimulus with respect to reference directions (Luu and Stocker, 2018). We found that while these models did produce repulsion effects, they could not account for the magnitude of the effects observed in our data and thus did not provide a good fit (see **Section A.2 in Appendix**), suggesting that more research is needed for understanding the mechanisms of reference repulsion effects.

To further investigate the role of repulsive biases in our task, we constructed a model (BP-AR4) which was agnostic to the mechanisms producing the repulsive cardinal biases, but aimed to capture them explicitly. While this model did account for the biases with respect to the cardinals, these effects proved to not fully account for the non-zero biases at the most frequently presented motion directions. To fully account for the latter biases we had to assume an additional repulsive bias from the central reference angle - model BP-AR40. Note that the central reference angle is not cued explicitly and would have to be inferred based on the statistics of the presented motion directions. Apart from that, we

can only speculate what mechanisms are responsible for the emergence of such effects. Noteworthily, our model selection is based on fixed-effects model comparison, which favoured BP-AR40 model, while random-effects BMS analysis was less conclusive. Model recovery analysis also showed that the data generated by BP-AR40 model was more frequently better recovered by BP-AR4 model. Together, this suggests that while there is evidence on the group level for the repulsion from the central reference angle, for many participants this repulsion might be not strong enough to justify BP-AR40 model. However, performing the same analysis on an independent dataset from Chalk et al. (2010), we again found BP-AR40 model to be the winning model, which provided additional confidence in the result that repulsion from the central reference angle has a significant effect in healthy controls (see **Section A.3 in Appendix**).

We found that BP-AR40 preserved the negative relationship between AQ and sensory uncertainty found in the previous analyses, confirming that this relationship was not confounded by the hitherto unaccounted repulsive biases. However, BP-AR40 results also suggested some additional effects to be present. AQ was found to be negatively correlated with prior-based lapse rate ($\tau_b = -0.155, p = .041$) - which had not reached significance in the previous BAYES_P model analysis ($p = 0.083$; see **Supplementary Fig. 6** in **Chapter 4**). Reduced prior-based lapses would be consistent with the reduced number of hallucinations found along AQ - which is the same relationship that was found in patients with SCZ (in **Chapter 3**).

Intriguingly, along schizotypy traits - that had shown no effects in the previous BAYES_P analysis - we found the repulsion from the central reference angle to be reduced. This relationship was found using both RISC and SPQ measures. In line with this, we found that SCZ group was best fit by BP-AR4 model, meaning

that repulsion from the central reference angle was negligible in this group. Interestingly, BP-AR4 model was also found to be the winning model in ASD group. Together, this presents an interesting result that might relate to transdiagnostic features of ASD and SCZ and which also extends to schizoptypy traits in the general population. The meaning and significance of this finding, however, is difficult to gauge. Not only are the mechanisms producing repulsive biases not well understood (e.g, see Kuang, 2019), repulsion from an arbitrary implicit reference direction (in our case, the central reference angle), has not been reported before. Finally, it is important to emphasize that the effects of repulsion from the central reference angle are quite subtle to begin with, as evidenced by the analysis of an independent dataset from Chalk et al. (2010), where each participant had ~ 3 times as much data, but random-effects BMS was still inconclusive (see **Section A.3 in Appendix**).

6

Within-trial dynamics in the Moving Dots task

6.1 Introduction

In all of our analysis thus far we have been treating each trial as a durationless instant, ignoring the duration of exposure to stimulus. In our task, exposure duration is controlled by reaction times - a trial terminates once a response is made (within a window of 3 seconds). Intuitively, we might expect that the more time is spent accumulating sensory evidence, the less biases we would see. For instance, in the context of visual illusions - which are typical examples of perceptual Bayesian inference - it has been shown that illusion strength varies as a function of exposure duration (Calvert and Harris, 1988) and might mediate individual differences in susceptibility to illusions (Bressan and Kramer, 2013). Importantly, in our data, we find SCZ group to be slower and ASD group to be faster than controls, while showing no differences in biases of motion direction estimation. Thus, the differences in reaction times might be a potential confound for the lack of differences in biases. This warrants a reanalysis of task performance within the framework of within-trial evidence accumulation. Interestingly,

the effects of within-trial dynamics on the observed Bayesian biases have not been accounted for in any of the recent studies investigating SCZ and ASD (see studies summarized in **Table 1.1** and **Table 1.2**), which might be a factor contributing to the mixed findings. More generally, experimental work exploring the link between within-trial evidence accumulation and Bayesian biases is sparse, motivating a more general analysis of how the effects observed within each of these frameworks relate to each other.

Accumulation of noisy sensory evidence has been most widely studied using 2-alternative forced choice (2AFC) tasks, modelling this process as a directed random walk to a bound - Drift-Diffusion Model (DDM; Ratcliff, 1978; for a review see Ratcliff and McKoon, 2008). DDM has been shown to be an optimal mechanism for 2AFC decision making (e.g., Bogacz et al., 2006), and has been directly linked to Bayesian inference (Bitzer et al., 2014; Fard et al., 2017). In DDM the decision-making process is summarized by: 1) *drift rate*, which captures how quickly the sensory evidence is accumulated, 2) *decision boundary*, which captures how much sensory evidence is needed for a choice to be made, and 3) *non-decision time*, which captures an additional time that is required for neural encoding and motor response. When there is bias or preference for one of the choices, this can be accounted for by including a *starting point* parameter, which corresponds to a displacement towards one of the boundaries at the beginning of a trial. DDM has been successfully applied for studying decision making in a whole range of different tasks, from simple motion perception, to memory retrieval, to lexical decision tasks (e.g., see Voss et al., 2013; Ratcliff and McKoon, 2008, for reviews). Furthermore, the latent variables captured by the model parameters have been associated with distinct brain networks in fMRI human studies (see Mulder et al., 2014, for a review), while the sensory accumulation process itself has been linked to the build-up of activity in single-cell recordings in monkeys and rats (Hanes

and Schall, 1996; for a review see Smith and Ratcliff, 2004).

Various task conditions and individual differences have been associated with different DDM parameters. Drift rate has been shown to have an inverse relationship with task difficulty (Voss et al., 2004) and a direct relationship with fluid intelligence and working memory capacity (Schmiedek et al., 2007), as well as motivation (Spaniol et al., 2011). In our task, difficulty is dependent on sensory uncertainty, which depends on stimulus contrast levels. Thus, we would expect drift rate to be directly related to sensory precision in a Bayesian model. Decision boundary is considered to represent caution and has been shown to increase with age (Ratcliff et al., 2010) and in response to task instructions that emphasize accuracy over speed (Voss et al., 2004). Non-decision time has also been positively associated with age (e.g., Ratcliff et al., 2006) and negatively associated with task readiness (Schmitz and Voss, 2012). Finally, starting point has been shown to capture bias towards more rewarding or more likely outcomes (Mulder et al., 2012; Summerfield and Koechlin, 2010; Voss et al., 2004). Thus, we would expect starting point to relate to prior precision in a Bayesian model.

In schizophrenia (SCZ), despite the slower processing speed and decreased accuracy being well-documented (e.g., Schatz, 1998), very few studies have investigated these effects within the DDM framework. Heathcote et al. (2015) studied patients with schizophrenia in left-vs-right motion discrimination task with random-dot kinematogram (RDK) stimuli and reported larger decision boundaries compared to controls. This was interpreted as a compensatory strategy to reduce inaccuracy arising from poorer sensory sensitivity. In a probabilistic reward and punishment learning task, Moustafa et al. (2015) also found SCZ to exhibit larger decision boundaries, but also longer non-decision times. Mathias et al. (2017) tested patients with schizophrenia in a processing-speed digit-symbol

coding task and found no differences in decision boundaries, but patients showed increased non-decision time and decreased drift rate.

In autism spectrum disorders (ASD), alterations in the temporal features of sensory processing have been found across multiple tasks (see Robertson and Baron-Cohen, 2017, for a review), but DDM has not been applied to study these effects until recently. In an orientation discrimination task with Gabor patches, Pirrone et al. (2017) found adults with ASD to exhibit increased decision boundary and increased non-decision time, while drift rate was comparable to that of controls. However, these effects were not found to extend to autistic traits in the general population (Pirrone et al., 2018). Another study (Karalunas et al., 2018) reported children with ASD to have increased decision boundary and decreased drift rate on go trials from the Stop Signal task. Finally, Powell et al. (2019) applied DDM to study face recognition and gaze discrimination in adolescents with ASD and reported reduced decision boundary and reduced drift rate across both tasks, but only when excluding participants with IQ below 85 (when including all participants, reduced decision boundary was found only for face recognition, while reduced drift rate was found only for gaze discrimination). While these findings conflict with the other two studies, it could be reconciled by the fact that here the stimuli were social and relatively complex, while Pirrone et al. (2017) and Karalunas et al. (2018) used simple stimuli; people with ASD are known to exhibit differential performance depending on the complexity of stimulus (Bertone et al., 2005). Furthermore, Karalunas et al. (2018) used a task that involves response inhibition, which might also be capturing effects that are specific to this domain.

To add to this body of research we investigated the effects of within-trial dynamics in the Moving Dots task, which we hitherto have largely ignored. We

wanted to know how within-trial dynamics were related to the strength of the Bayesian effects as well as to the individual and group differences found in our previous analysis. Unlike the studies discussed above, which used 2AFC tasks, in our task responses were not on a binary but on an interval scale. While in recent years various extensions of DDM have been proposed, including DDMs for multiple choice tasks (see Forstmann et al., 2016, for a review), full DDMs for continuous scale responses have not been introduced (although see Smith, 2016, for an analytic derivation of continuous response DDM for tasks with no bias). We therefore constructed and validated Continuous Choice Drift-Diffusion Model (CDM), which can incorporate bias terms of various complexity (including non-unimodal priors, such as bimodal prior used in our task) as well as trial-to-trial variability in any of the model parameter values.

6.2 Methods

6.2.1 Behavioral data analysis

To assess the association between reaction time and other performance measures in our task, we analyzed both within-subject and between-subject reaction time (RT) effects. Both of these effects were tested by fitting a linear mixed model in SPSS version 25. For the between-subject effects, stimulus direction was included as a repeated-measures factor and the mean RT at each stimulus direction was included as a covariate. We tested for the main effects of stimulus direction and RT as well as their interaction. Note that for illustration purposes we separated individuals into 'fast' and 'slow' groups by ordering all individuals based on their median RT and selecting 30 fastest and 30 slowest individuals. The linear mixed model described above, however, was fit to all individuals.

For analysing within-subject effects, we obtained 'fast' and 'slow' trials by

splitting each participant's trials based on their median reaction time and removing 30% of the trials around the median to maximize the differences between 'fast' and 'slow'. In the linear mixed model, this was included as a repeated-measures factor that had two values ('fast'/'slow'). Together, this model included the main effects of both stimulus direction and RT category as well as their interaction.

6.2.2 Models

6.2.2.1 Continuous Choice Drift-Diffusion model: CDM

To construct Continuous Choice Drift-Diffusion Model (CDM), the classical DDM had to be extended from one space dimension to two. In addition, the model had to capture the bimodality of the prior acquired in our task - this was achieved by having two particles drifting simultaneously (**Fig. 6.1**):

$$\vec{X}_n(t+dt) = \vec{X}_n(t) + v \cdot dt \cdot \hat{u}(\theta_{act}, \vec{X}_n(t)) + \vec{\eta} \tag{6.1}$$

$$\vec{X}_p(t+dt) = \vec{X}_p(t) + v \cdot dt \cdot \hat{u}(\theta_{act}, \vec{X}_p(t)) + \vec{\eta} \tag{6.2}$$

where $\vec{X}_n = [x_n, y_n]$ and $\vec{X}_p = [x_p, y_p]$ are the coordinates of *negative* and *positive* particles, respectively, t is time, dt is a time step, v is the mean drift rate, and $\vec{\eta}$ is random Gaussian noise:

$$\vec{\eta} \sim N_2(0, \vec{\Sigma}); \quad \vec{\Sigma} = [s \cdot \sqrt{dt}, s \cdot \sqrt{dt}] \quad , \tag{6.3}$$

where s is a scaling parameter for the diffusion process; it was set to 0.1 by convention.

$\hat{u} = [u_x, u_y]$ in Eq. 6.1 and 6.2 is a unit vector of drifting direction with respect to the current position of a particle, $\vec{X}(t) = [x(t), y(t)]$, and depending on the stimulus motion direction, θ_{act}; $\hat{u}$ was computed via:

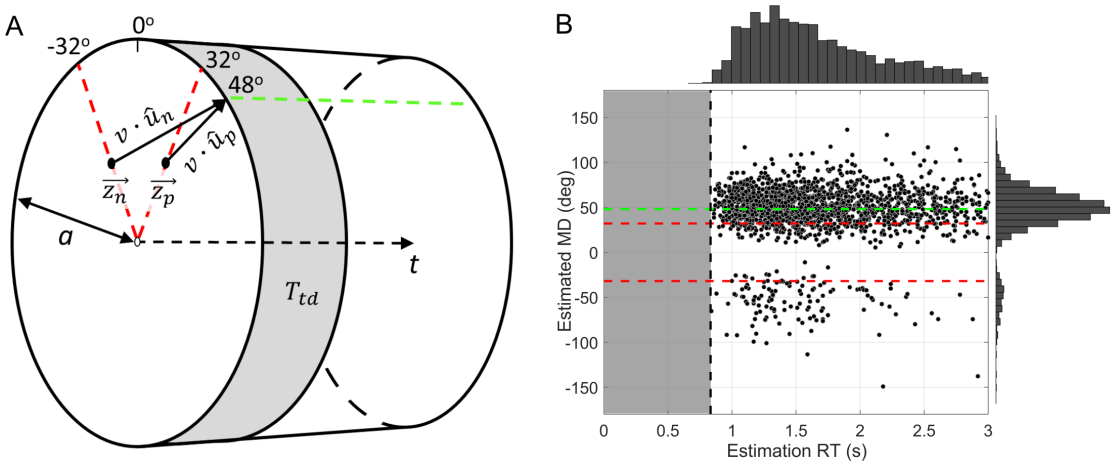


Figure 6.1: CDM schema. (**A**) A cylinder illustrating CDM with a stimulus motion direction $\theta_{act} = 48°$. On a given trial, two particles drift in two space dimensions until one of them reaches the circular boundary (or until 3 seconds elapse). a - the radius of the circular boundary, $\vec{z}_n$ and $\vec{z}_p$ - the starting points of the two particles, v - the mean drift rate, $\hat{u}_n$ and $\hat{u}_p$ - unit vectors of the drifting direction for each particle (see Eq. 6.4 and 6.5), T_{nd} - non-decision time (the shaded area); t denotes the time dimension indicated with the dashed arrow. (**B**) Flattened out view of the cylinder showing the distribution of reaction times and estimated motion directions produced by running 2,000 simulations. The red dashed lines indicate the most frequently presented motion directions ($\pm 32°$), while the green dashed line indicates stimulus motion direction on the current trial ($48°$).

$$\theta_{drift} = \text{atan2}(a \cdot \sin(\theta_{act}) - y(t),\ a \cdot \cos(\theta_{act}) - x(t)) \tag{6.4}$$

$$\hat{u} = [\cos(\theta_{drift}),\ \sin(\theta_{drift})] \tag{6.5}$$

where θ_{drift} is the direction of drift in radians and a is the distance from the origin to the boundary - i.e. the radius of the cylinder.

The locations of the particles at the beginning of a trial are determined by the starting points:

$$\vec{X}_n(t=0) = \vec{z}_n;\quad \vec{X}_p(t=0) = \vec{z}_p \tag{6.6}$$

In line with our previous modelling, the starting points $\vec{z}_n$ and $\vec{z}_p$ were con-

strained to be symmetrical around the central reference angle such that:

$$\vec{z}_n = [z_d,\ -z_\theta];\quad \vec{z}_p = [z_d,\ z_\theta]\quad , \tag{6.7}$$

where z_d is the distance from the origin and z_θ is the direction with respect to the origin.

The stimulus encoding time (a brief period between the start of a trial and the beginning of the drift-diffusion process) and motor response time (the time taken to make a response after the drift-diffusion process ends) were accounted for by including a non-decision time parameter, T_{nd}, which captured the combined duration of the two aforementioned processes. Incorporating T_{nd}, the reaction time then becomes:

$$RT = t + T_{nd} \tag{6.8}$$

where t is the time elapsed in the drift-diffusion process.

With the time step $dt = 0.01$ the locations of the particles were recursively updated via Eq. 6.1 and 6.2 until 3 seconds had elapsed ($RT \geqslant 3$) or the decision boundary was reached by one of the particles: $|\vec{X}_n(t)| \geqslant a$ or $|\vec{X}_p(t)| \geqslant a$. The location $[x_a, y_a]$ at which a particle touched the boundary corresponded to the perceived motion direction, θ_{perc}:

$$\theta_{perc} = atan2(y_a, x_a)\quad . \tag{6.9}$$

The estimated motion direction, θ_{est}, can then be obtained by accounting for motor noise:

$$p(\theta_{est}|\theta_{perc}) = V(\theta_{perc}, \sigma_m) \quad . \tag{6.10}$$

The joint probability distribution of reaction times and estimated motion directions, for a given motion direction, $p_k(RT, \theta_{est}|\theta_{act})$, was obtained by running 2,000 simulations. To minimize the effects of random fluctuations in this distribution, instead of building a histogram, we computed bivariate kernel density estimate using *ksdensity* function in Matlab. Finally, to account for a possibility of random responses, we introduced an additional parameter α:

$$p(RT, \theta_{est}|\theta_{act}) = (1-\alpha) \cdot p_k(RT, \theta_{est}|\theta_{act}) + \alpha \cdot U(RT, \theta_{est}) \quad , \tag{6.11}$$

where $U(RT, \theta_{est})$ is a bivariate uniform distribution, reflecting the assumption that lapse responses are distributed uniformly in time and direction.

6.2.2.2 Variants of CDM

In the standard DDM model, across-trial variability in model parameter values is often found it to be necessary to account for features in empirical data (e.g., Ratcliff, 2013; Ratcliff et al., 2016). Thus, we also considered more complex variants of CDM model, by allowing for trial-to-trial variability in model parameters. We reasoned that there might be a considerable variation in the drift rate (*CDMv*) or non-decision time (*CDMt*) due to the variation in stimulus contrast (arising from contrast staircases). We also considered that prior effects might not be completely fixed throughout the task, and therefore we constructed a model where the distance of the starting points, z_d, varied from trial to trial (*CDMz*). Finally, we also built a model to test variability in the decision boundary (*CDMa*) to account for a possibility of changes in speed-accuracy trade-off strategy during the task.

All of these models were implemented by introducing an additional free parameter (σ_v, σ_t, σ_z or σ_a for models CDMv, CDMt, CDMz and CDMa, respectively), which captured the standard deviation from the mean value of a given parameter. Using these summary statistics, a new value of a given parameter was sampled for each of the 2,000 trials in the model simulation.

6.2.3 Model fitting

The models were fit by maximizing the log likelihood function for the joint probability of motion estimation and reaction time data for each participant individually. The maximum likelihood was found using *fminsearchbnd* function in Matlab, by minimizing negative log-likelihood. First we performed a wide exploration of the parameter space to determine the full range of values that each of the parameters can take. For CDM this established the following range of values: $a \in [0.2,\ 0.7]$, $v \in [0.05,\ 0.3]$, $z_d \in [0,\ a/1.2]$, $z_\theta \in [0,\ 360]$, $T_{nd} \in [0,\ 1.2]$, $\alpha \in [0,\ 0.2]$. For other models, the range of each parameter was determined to be: $\sigma_a \in [a/10,\ a]$ for CDMa, $\sigma_v \in [v/10,\ v]$ for CDMv, $\sigma_T \in [T_{nd}/10,\ T_{nd}]$ for CDMt and $\sigma_z \in [z_d/10,\ z_d]$ for CDMz. Then, to reduce the possibility of convergence at local maxima, the main model fitting procedure was performed by running 20 different initializations with parameter values sampled randomly from the ranges defined above.

6.2.4 Model comparison

As in the previous chapters, to compute the log of model evidence we used Bayesian Information Criterion (BIC) approximation:

$$\log(P(D|M)) \approx -\frac{BIC}{2} = \log(P(D|M,\hat{\Theta})) + k \cdot \log(n) \tag{6.12}$$

where M is model, D is observed data and P (D|M,$\hat{\Theta}$) is the likelihood of

generating the experimental data given a set of parameters,$\hat{\Theta}$; k is the number of model parameters and n is the number of data points (i.e. trials). To compare the models we performed a random-effect Bayesian model selection (BMS) analysis (Stephan et al., 2009b) using VBA Matlab toolbox (Daunizeau et al., 2014). In random-effect BMS, model comparison is based on exceedance probability (EP), which expresses the likelihood that any given model is more frequent than all other models. Correcting for chance variations in model frequencies (Rigoux et al., 2014), we can obtain protected exceedance probability (PEP):

$$PEP_k = EP_k(1-BOR) + \frac{1}{K}BOR \quad , \tag{6.13}$$

where subscript k denotes kth model from the total of K models being compared; BOR is Bayesian Omnibus Risk, which quantifies the chance likelihood of the observed model frequencies:

$$BOR = \frac{p(y|H_0)}{p(y|H_0)+p(y|H_1)} \quad , \tag{6.14}$$

where y contains log model evidences of all subjects and all models, H_0 assumes that model frequencies are all the same and H_1 assumes that at least one of the models is more frequent. Thus, the higher the BOR estimate, the less differences there will be in PEP of different models.

As random-effects BMS was not always conclusive, we also performed fixed-effects model comparison. For this we summed the BIC across all participants and used this sum to compare the models. Unlike random-effects model comparison, this assumes that all participants in the group are best described by the same model.

6.2.5 Model and parameter recovery

Modelling results, and their robustness, can be affected by many factors, such as parameterisation of models, amount of data used for model fitting, chosen model fitting methods, etc. A general way to account for these factors and to assess the overall reliability of modelling results is to perform model and parameter recovery (e.g., see Palminteri et al., 2017). Given the constraints of the mentioned factors, model recovery shows the distinguishability of models, while parameter recovery shows reliability of parameter estimates.

For each model, we generated 100 sets of parameters (100 synthetic individuals), by sampling the distributions of parameter values found in our general population sample. We then simulated 200 responses for each synthetic participant (just as we had ∼200 responses per participant in our data). Every model was then fitted to every simulated dataset following the same model fitting procedure as described in the previous section. We then performed model comparison for every simulated dataset, also as described in the previous section.

We also analysed how well the parameters were recovered for the winning model. This was done by comparing parameters that simulated the data ('actual parameters') with the parameters estimated from the simulated data ('recovered parameters') using the coefficient of determination (R^2) obtained from *regress* function in Matlab.

6.2.6 Datasets

The behavioral data analysis of within- and between-subject effects, as well as most of the modelling analysis was done using our largest general population data sample ($N = 83$; **Chapter 2**). Then we used the same modelling approach to investigate SCZ (from **Chapter 3**) and ASD sample (from **Chapter 4**).

6.2.7 Statistical tests

Due to outliers and non-normalities being common in our data, we used non-parametric tests as a default in our analysis. For correlation we used Kendall's Tau-b, while for group differences we used Mann-Whitney U test. Wherever relevant, the evidence for the null hypothesis was evaluated by computing Bayes factors (BF_{01}) using JASP version 0.10 (JASP Team, 2019), with the default Cauchy prior of scale 0.707. Bayes factor below 3 is generally considered inconclusive, above 3 is considered moderate evidence, above 10 is strong evidence and above 30 is very trong evidence (Kass and Raftery, 1995).

6.3 Results

6.3.1 Behavioral Results

To assess the associations between reaction time and other performance measures in our task, we investigated both between-subject and within-subject effects (**Fig. 6.2**). Analyzing reaction time effects across participants (between-subject effects), faster individuals were found to be associated with more lapse estimations (**Fig. 6.2C**; $F(1,605.8) = 6.9,\ p = .009$), more estimation variability (**Fig. 6.2B**; $F(1,631.2) = 15.9,\ p < .001$) and more positive bias (**Fig. 6.2A**; $F(1,516.2) = 5.1,\ p = .025$). The latter finding is quite unexpected and suggests faster individuals to exhibit a clockwise bias. No interaction effects between RT and motion direction were found for any of the above measures (bias: $F(8,153.1) = 0.5,\ p = .819$; variability: $F(8,132.7) = 0.6,\ p = .796$; lapses: $F(8,139.4) = 0.6,\ p = .738$).

Examining reaction time effects across trials (within-subject effects), we found faster responses to be associated with increased bias (**Fig. 6.2E**; motion di-

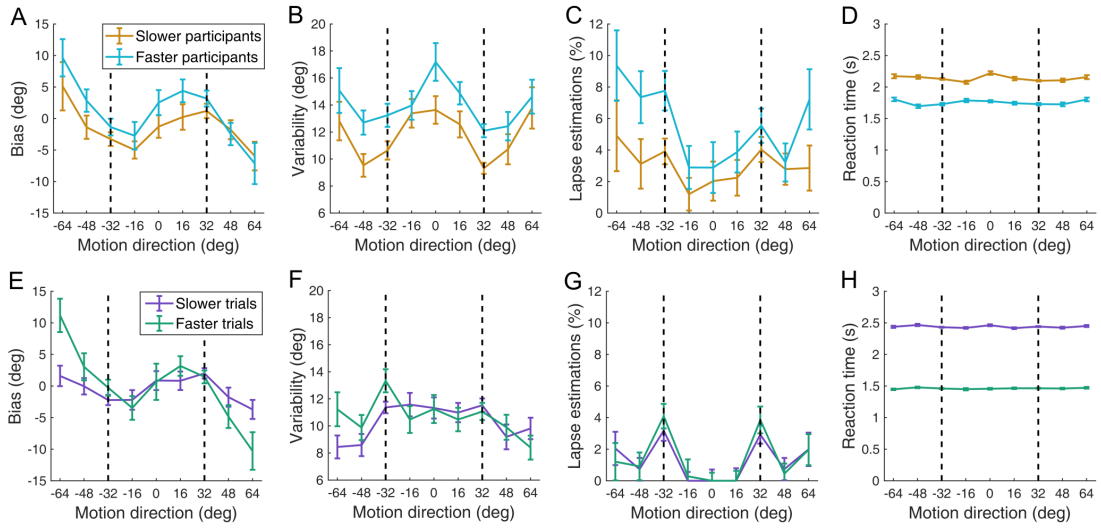


Figure 6.2: The relationships between reaction time and estimation performance measures. (**A-D**) between-subject effects: fast vs slow participants (30 fastest vs 30 slowest participants), (**E-H**) within-subject effects: fast vs slow trials (35% fastest and 35% slowest trials for each individual).(**A,E**) Estimation bias, (**B,F**) estimation variability, (**C,G**) lapse estimations, (**D,H**) reaction time. The vertical dashed lines correspond to the two most frequently presented motion directions ($\pm 32^{\circ}$). Error bars represent within-subject standard error.

rection*RT: $F(8, 269.3) = 3.1,\ p = .002$), while there were no clockwise or anticlockwise biases (main RT effect: $F(1, 803.1) = 0.6,\ p = .449$). Faster responses were also not associated with any differences in response variability (**Fig. 6.2F**; main effect: $F(1, 1299.4) = 1.7,\ p = .196$; motion direction*RT: $F(8, 286.3) = 1.1,\ p = .399$) or lapse estimations (**Fig. 6.2G**; main effect: $F(1, 1163.1) = 0.4,\ p = .551$; motion direction*RT: $F(8, 286.7) = 0.5,\ p = .827$).

6.3.2 Modelling results

To better understand the within-trial dynamics in our task, we fitted and compared five different CDM variants. We computed log model evidence using BIC approximation and used random-effects Bayesian model selection analysis to compare the models. We found the simplest model CDM, which did not assume any trial-to-trial variability in model parameters, to be the best model with protected exceedance probability of 1 (**Fig. 6.3A**). In the context of within-subject ef-

fects discussed in the previous section, this means that faster and more biased responses on some of the trials did not arise from trial-to-trial parameter variability, but simply arose from noise due to chance. In other words, on trials where noise coincided with the direction of expectations, this lead to faster and more biased responses. The reaction time within-subject effects on bias can also be explicitly shown to be predicted by the model (see **Section B.1 in Appendix**).

CDM provided a good fit to the data both on the individual level (**Fig. 6.3B-J**) and on the group level (**Fig. 6.3K-N**). On the individual level, the most notable feature that the model captured was the prior-based lapses: the occasional estimations around the most frequent directions even when the presented motion direction was far away from it (**Fig. 6.3B-J**). It is worth noting that prior-based lapses were not included in the model explicitly, but arose naturally from having two drifting particles representing the acquired bimodal prior. On the group level, CDM captured the characteristic shape of the bias curve (**Fig. 6.3K**), reduced variability and reaction time around the most frequent directions (**Fig. 6.3L,N**), and the level of lapse estimations (**Fig. 6.3M**). The only notable discrepancy is the overestimation of variability by CDM (**Fig. 6.3L**). The source of this overestimation remains unclear, but it is interesting to note that the Bayesian models analyzed in the previous chapters exhibited similar overestimation of variability (e.g., see **Fig. 6B,F** in **Chapter 3**).

6.3.2.1 Model and parameter recovery results

To further analyze the robustness of modelling results, we performed model and parameter recovery. We found that most models were recovered reasonably well (**Fig. 6.4**), with CDM and CDMv showing a perfect recovery, while CDMa and CDMt were recovered half the time, with CDM providing a better fit for the other half. Finally, CDMz showed the poorest recovery with CDM providing

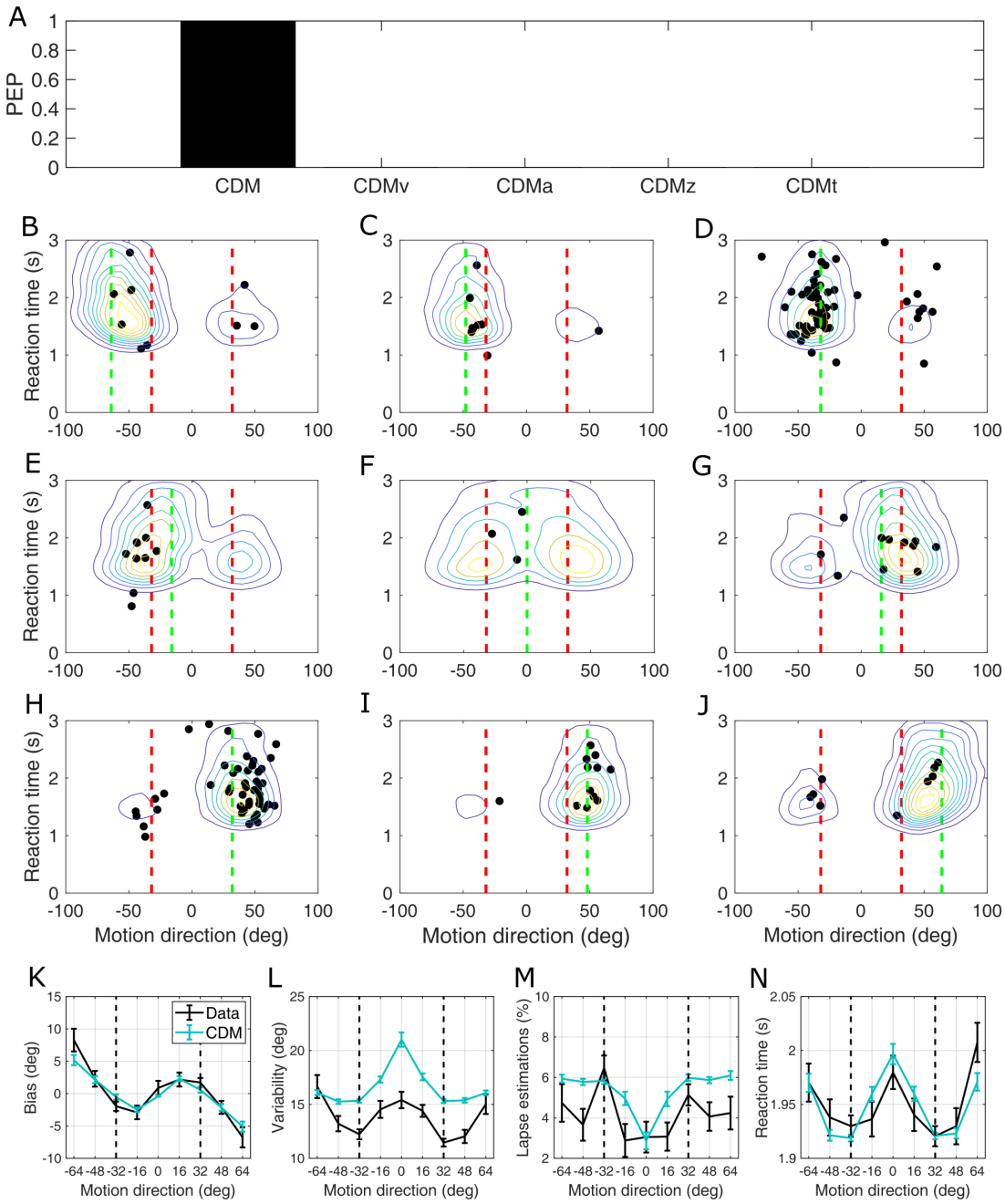


Figure 6.3: Modelling results and CDM fits to the data. (**A**) Random-effects Bayesian model selection showing CDM to be the most likely model with protected exceedance probability (PEP) equal to 1. (**B-J**) CDM fits across different stimulus motion directions for a single individual. Black data points represent behavioral data. Contour lines denote joint probability distributions predicted by the model. Green dashed lines indicate presented motion direction (in the presented order: $-64°$, $-48°$, $-32°$, $-16°$, $0°$, $16°$, $32°$, $48°$, $64°$). Red dashed lines correspond to the most frequently presented directions ($\pm 32°$). (**K-N**) CDM fits averaged across all participants and across all responses for a given motion direction. (**K**) Estimation bias, (**L**) estimation variability, (**M**) lapse estimations, (**N**) estimation reaction time. The vertical dashed lines denote the two most frequently presented directions. Error bars represent within-subject standard error.

consistently better fit to the data generated by CDMz. Non-perfect model recovery for models with the trial-to-trial variability parameters is not too surprising. We have to keep in mind that CDM was nested in all other models and that the variability parameter for each model was based on what was empirically found by fitting that model to behavioral data. Thus, if participants did not exhibit large trial-to-trial variability to begin with, these effects would also be small in the simulated data, thus allowing CDM to provide a more parsimonious explanation of the simulated data.

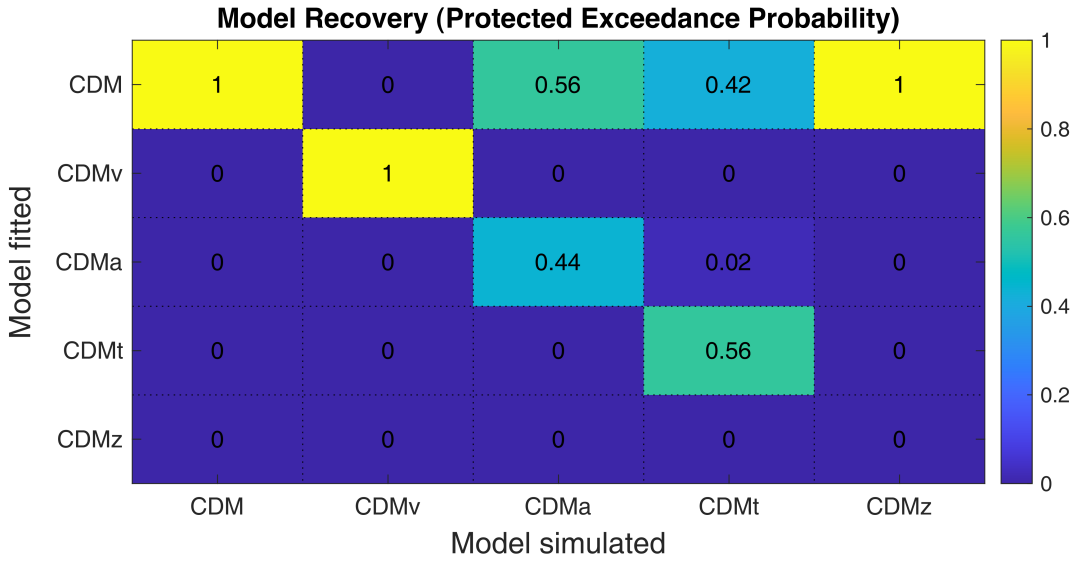


Figure 6.4: Model recovery based on protected exceedance probability. X-axis - models that simulated the data, Y-axis - models that were fitted to the simulated data.

Another possibility is that we did not have enough data to capture trial-to-trial variability in model parameters. The classical one-dimensional DDM requires ∼250 to reliably capture trial-to-trial variability in model parameters (e.g., Ratcliff and Tuerlinckx, 2002), while our model was two-dimensional and we had ∼ 200 trials per participant. To examine the possibility that more data would reveal a different winning model, we performed model selection on an independent dataset from Chalk et al. (2010), where each participant had ∼700

trials of data. CDM was still found to be the winning model (see **Section B.3 in Appendix**).

For the winning model, CDM, we found that most parameters were recovered well (**Fig. 6.5**) with $R^2 = 0.80$ for drift rate v, $R^2 = 0.67$ for non-decision time T_{nd}, $R^2 = 0.66$ for lapse estimations α, $R^2 = 0.53$ for starting point distance z_d, $R^2 = 0.44$ for decision boundary a; starting point orientation, z_θ, showed the worst parameter recovery with $R^2 = 0.31$, but this was also affected by a few outliers (removing outliers that were more than 3 median absolute deviations away from the median resulted in $R^2 = 0.40$).

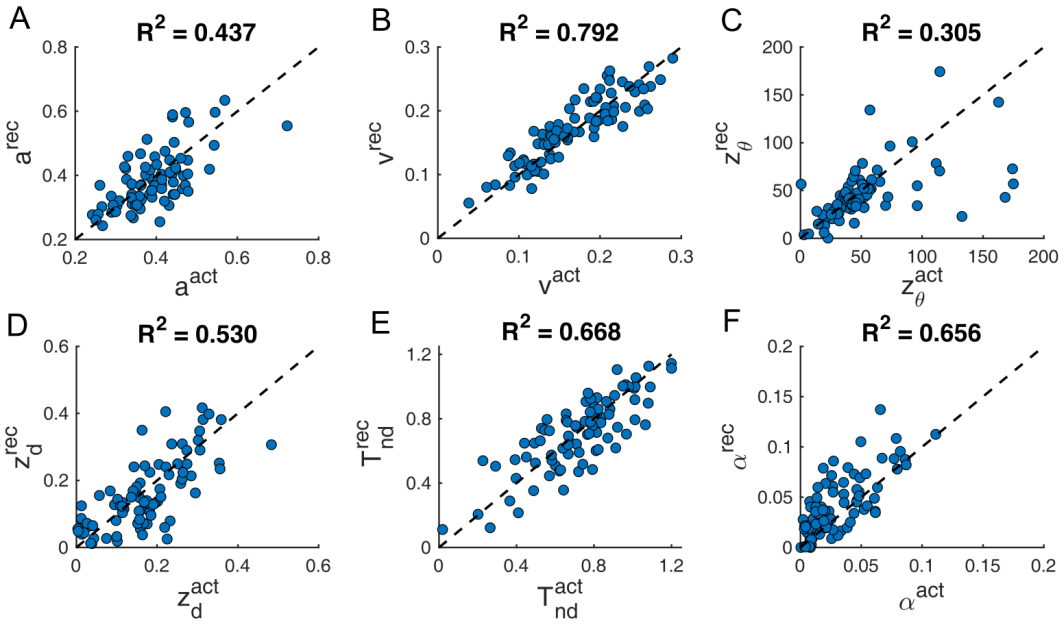


Figure 6.5: Parameter recovery for CDM. (**A**) a - decision boundary, (**B**) v - mean drift rate, (**C**) z_θ - starting point angle, (**D**) z_d - starting point distance, (**E**) T_{nd} - non-decision time, (**F**) α - random lapse responses. X-axes – actual parameters used for simulating the data (denoted with the superscript 'act'); Y-axes – recovered parameters (denoted with the superscript 'rec') after fitting the model to the simulated data. The dashed diagonal line is a reference line indicating perfect parameter recovery. The coefficient of determination is reported above each panel.

6.3.2.2 CDM simulations

To gain intuition about the effects of different parameters on task performance, we ran model simulations by manipulating each parameter separately. We used the average parameter values found in our sample as a baseline ($a = 0.48$, $v = 0.19$, $z_d = 0.28$, $T_{nd} = 0.8$, $\alpha = 0.06$) and simulated data by increasing and reducing the value of each model parameter by 30% (**Fig. 6.6**). Decision boundary and drift rate were both negatively associated with the strength of interaction effects (between a given measure and presented motion direction) for bias, variability and lapse estimations, but had different effects on reaction time, with larger decision boundary leading to longer mean reaction times, and larger drift rate leading to shorter mean reaction times. It is also interesting to note that compared to decision boundary, drift rate had smaller effect on reaction times, while for the other measures the effects were comparable. Starting point distance was positively associated with the strength of interaction effects for all of the measures and negatively associated with mean reaction times. Finally, non-decision time only affected mean reaction times, while random lapse parameter only affected lapse estimations.

6.3.2.3 CDM parameters vs personality traits

Next, we studied how variation in CDM model parameters related to autistic (AQ) and schizotypy (SPQ and RISC) traits. AQ scores were found to positively correlate with mean drift rate ($\tau_b = 0.157, p = .039$; **Fig. 6.7B**), suggesting that increased sensory precision found in BAYES_P model might be underlied by faster sensory accumulation. No other parameters were influenced by AQ (**Fig. 6.7**): decision boundary a ($\tau_b = -0.004,\ BF_{01} = 7.0$), starting point direction z_θ ($\tau_b = 0.108,\ BF_{01} = 2.5$), starting point distance z_d ($\tau_b = -0.040,\ BF_{01} = 6.1$; starting point distance relative to the boundary, z_d/a : $\tau_b = -0.084,\ BF_{01} = 3.8$), non-decision time T_{nd} ($\tau_b = 0.040,\ BF_{01} = 6.1$), lapse

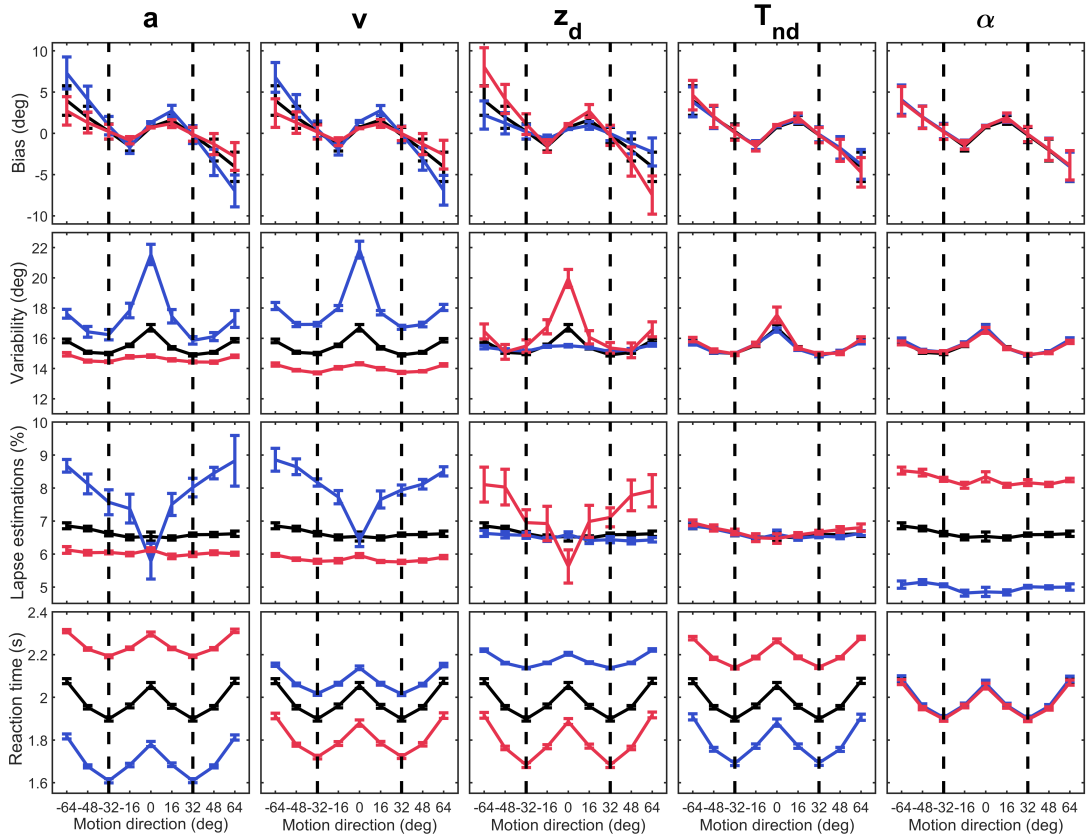


Figure 6.6: CDM simulations with different parameter settings. Black line - baseline performance ($a = 0.48$, $v = 0.19$, $z_d = 0.28$, $T_{nd} = 0.8$, $\alpha = 0.06$), blue line - 30% decrease in a given parameter value, red line - 30% increase in a given parameter value. Each column corresponds to manipulation of a different parameter (indicated at the top). a - decision boundary, v - mean drift rate, z_d - starting point distance, T_{nd} - non-decision time, α - random lapse responses. The vertical dashed lines denote the two most frequently presented directions.

estimations α ($\tau_b = -0.044$, $BF_{01} = 5.9$).

Schizotypy traits were found to be not associated with any of the CDM parameters: decision boundary a (RISC: $\tau_b = 0.018$, $BF_{01} = 6.8$; SPQ: $\tau_b = 0.136$, $BF_{01} = 2.3$), mean drift rate v (RISC: $\tau_b = -0.055$, $BF_{01} = 5.4$; SPQ: $\tau_b = 0.082$, $BF_{01} = 3.7$), starting point direction z_θ (RISC: $\tau_b = -0.149$, $BF_{01} = 1.0$; SPQ: $\tau_b = -0.033$, $BF_{01} = 4.6$), starting point distance z_d (RISC: $\tau_b = 0.044$, $BF_{01} = 5.8$; SPQ: $\tau_b = -0.022$, $BF_{01} = 4.7$), non-decision time T_{nd} (RISC: $\tau_b = -0.142$, $BF_{01} = 1.2$; SPQ: $\tau_b = 0.025$, $BF_{01} = 4.7$), lapse estimations α (RISC: $\tau_b = 0.083$, $BF_{01} = 3.8$; SPQ: $\tau_b = 0.063$, $BF_{01} = 4.1$).

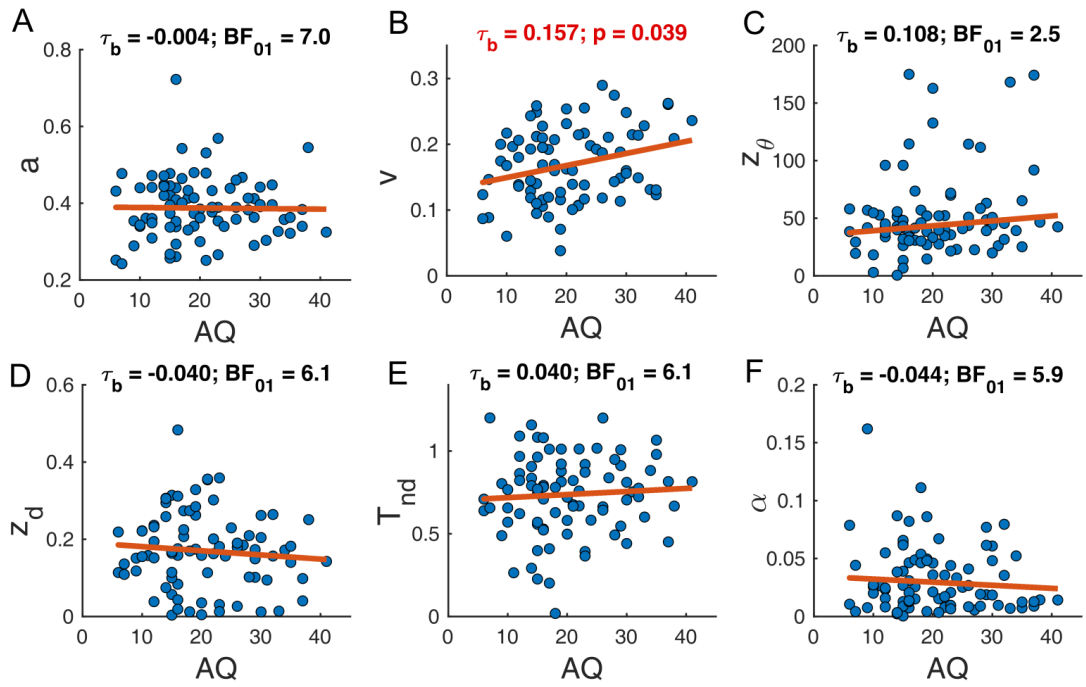


Figure 6.7: CDM parameters vs AQ traits. (**A**) a - decision boundary, (**B**) v - mean drift rate, (**C**) z_θ - starting point angle, (**D**) z_d - starting point distance, (**E**) T_{nd} - non-decision time, (**F**) α - random lapse responses. Two-tailed Kendal's Tau-b correlation results are reported above each panel. For non-significant correlations Bayesian Factors for the null hypothesis (BF_{01}) are presented.

6.3.2.4 Modelling results in SCZ and ASD groups

To investigate group differences in within-trial dynamics we performed the same modelling analysis in SCZ (described in **Chapter 3**) and ASD (described in **Chapter 4**) groups. Random-effects BMS and fixed-effects model comparison both indicated CDM to be the winning model in all of the groups **Fig. 6.8**): PEP = 0.998 and 67 BIC units lead in CTR, PEP = 0.832 and 15 BIC units lead in SCZ; in ASD group, while less definitively, CDM was still the most favoured model with PEP = 0.354 and 11 BIC units lead. Note that anything above 10 BIC units difference is considered very strong, decisive evidence (Kass and Raftery, 1995).

Next, we compared the groups in terms of their CDM parameter values. In SCZ (**Fig. 6.9**), we found increased decision boundary ($p = .012$; **Fig. 6.9A**),

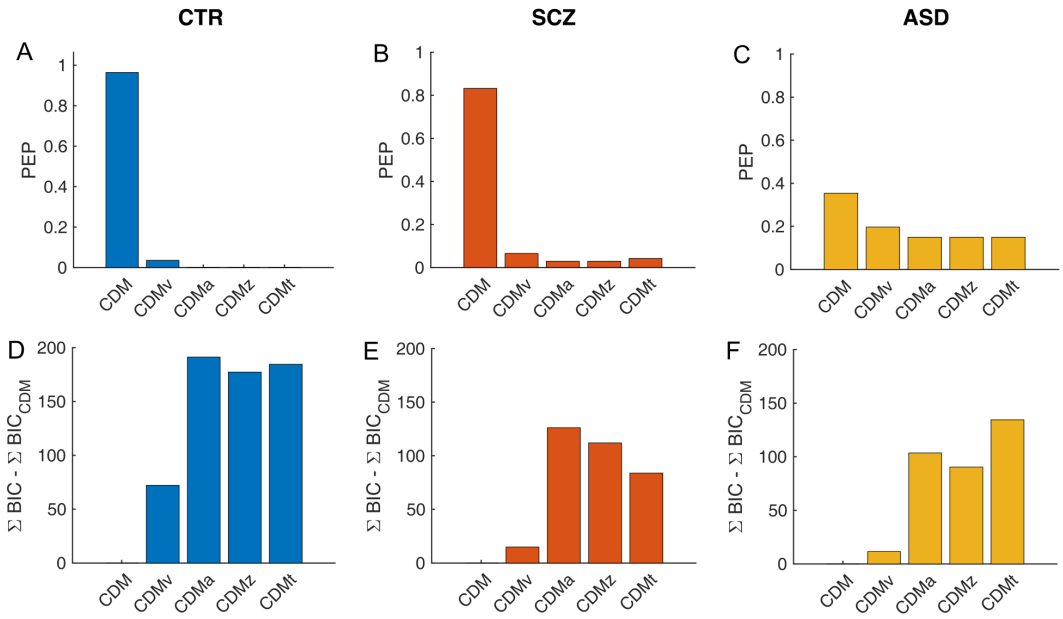


Figure 6.8: Model comparison for patient groups. (**A**-**C**) Random-effects Bayesian model selection; PEP - protected exceedance probability (**D**-**F**) Fixed-effects model comparison. Note that the plots show the difference between the model with the lowest BIC and the remaining models. (**A**,**D**) Controls, (**B**,**E**) participants with schizophrenia, (**C**,**F**) participants with ASD. Both approaches show CDM to be the winning model in each group. For model fits to the data see **Section B.4 in Appendix**.

which would explain why we see longer reaction times and less prior-based lapses in patients. Other model parameters showed no differences but were also inconclusive, with $BF_{01} < 3$ (**Fig. 6.9B-F**), except for starting point direction ($BF_{01} = 3.2$; **Fig. 6.9C**). Starting point distance relative to the boundary, z_d/a, was also not different in SCZ ($BF_{01} = 3.4$).

In ASD group, despite them exhibiting faster RTs in the behavioral data, we found no significant differences in any of the parameters (**Fig. 6.10**), including starting point distance relative to the boundary, z_d/a ($BF_{01} = 2.8$). However, some of the parameters did not have strong support for the null hypothesis either: decision boundary ($BF_{01} = 1.4$; **Fig. 6.10A**), starting point distance ($BF_{01} = 1.8$; **Fig. 6.10D**) and non-decision time ($BF_{01} = 2.7$; **Fig. 6.10E**), making them likely to be implicated in the observed RT differences.

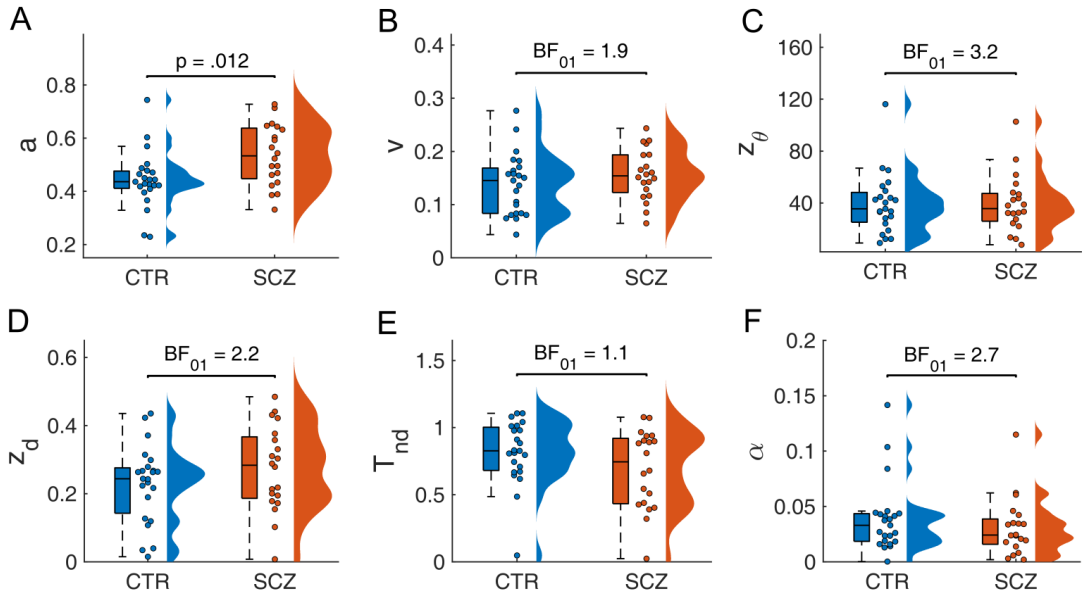


Figure 6.9: CDM parameter comparison between controls (CTR) and patients with schizophrenia (SCZ). (**A**) a - decision boundary, (**B**) v - mean drift rate, (**C**) z_θ - starting point angle, (**D**) z_d - starting point distance, (**E**) T_{nd} - non-decision time, (**F**) α - random lapse responses. Two-tailed Mann-Whitney U test results are reported with each panel. Significant differences are reported in p-values, while for non-significant differences Bayesian Factors for the null hypothesis (BF_{01}) are presented.

6.4 Discussion

In this chapter we sought to analyse the performance in the Moving Dots task within a framework that accounts for within-trial dynamics. To this end, we first investigated within- and between-subject effects of reaction time with respect to the performance measures of motion estimation. Secondly, to gain a more mechanistic insight into these effects, we constructed and analysed Continuous Choice Drift Diffusion Model (CDM).

We found that faster responses were associated with increased biases (within-subject effects). Model comparison, which showed CDM to be a better model than its variants that assumed trial-to-trial parameter variability, suggested that the observed within-subject effects were a consequence of random noise and chance. In other words, on trials when sensory noise happened to coincide with the di-

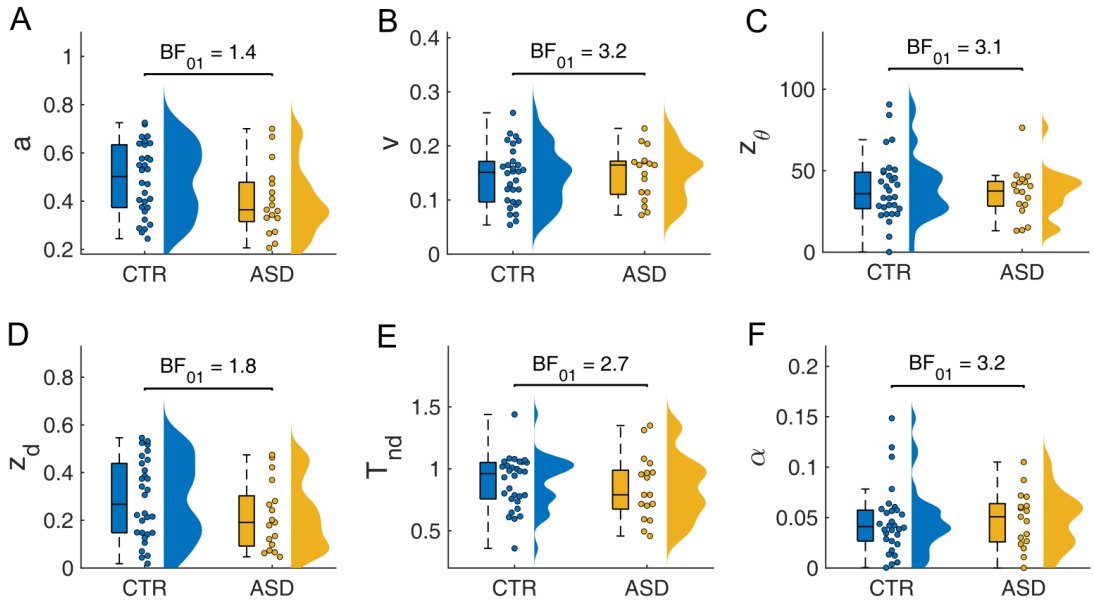


Figure 6.10: CDM parameter comparison between controls (CTR) and participants with autism spectrum disorder (ASD). (**A**) a - decision boundary, (**B**) v - mean drift rate, (**C**) z_θ - starting point angle, (**D**) z_d - starting point distance, (**E**) T_{nd} - non-decision time, (**F**) α - random lapse responses. Two-tailed Mann-Whitney U test results are reported with each panel. As no differences were statistically significant, only Bayesian Factors for the null hypothesis (BF_{01}) are shown.

rection of the prior expectations, this lead to faster responses and more bias.

Analysing between-subject effects, we found faster participants to exhibit more lapse responses and more variability. CDM parameter analysis showed these effects to be mediated by a combination of reduced decision boundary and increased random lapse responses (see **Section B.2 in Appendix**), suggesting the observed between-subject effects might have been driven by motivation (with less motivated participants trading off accuracy for speed and making more random responses). While faster individuals did not differ in the total amount of bias, the results showed an additional clockwise bias in their estimation responses. The latter result is rather surprising, but might relate to perceptual-motor clockwise/anti-clockwise bias reported in the literature (Karim et al., 2016). However, on the whole group level we observed no clockwise or

anti-clockwise bias.

Analysing CDM parameters with respect to autistic traits in the general population sample, we found higher AQ scores to be associated with faster drift rates, while for other parameters we found strong evidence for the null hypothesis. This means that the increased sensory precision in the previous Bayesian model analysis can be thought of as being underlied by faster accumulation of sensory evidence. Note that faster drift rate would also be consistent with reduced number of hallucinations and reduced prior-based lapses found along AQ after accounting for repulsive biases (in **Chapter 5**), even though the relationship with prior-based lapses was only trending in the original BAYES_P analysis (**Chapter 4**). Schizotypy traits were not associated with any of the CDM parameters, consistent with no behavioral effects.

Finally, we performed the same analysis on SCZ and ASD samples. Just as controls, both groups were best fitted by the simplest CDM model with no trial-to-trial variability in model parameter values. Parameter comparison showed SCZ group to have increased decision boundary, which explains longer reaction times and reduced hallucinations observed in this group. Most other parameter comparisons, while not significant, showed weak evidence for the null ($BF_{01} < 3$ for most parameters), not allowing to accept the null hypothesis. In ASD group we found no differences in any of the CDM parameters, which means that the observed faster reaction times were not underlied by a single parameter. Decision boundary, starting point distance and non-decision time, however, did not have strong evidence for the null ($BF_{01} = 1.4$, $BF_{01} = 1.8$ and $BF_{01} = 2.7$, respectively), and might have collectively contributed to faster performance. Unlike autistic traits in the general population, ASD group showed no increase in the drift rate ($BF_{01} = 3.2$).

Our finding of increased drift rate along AQ contrasts with the previous reports using DDM to study ASD. In an orientation discrimination task, Pirrone et al. (2017) reported no differences in drift rate, but increased decision boundary and increased non-decision time. This, however, was found in adults with ASD, while along autistic traits in the general population no effects were found at all (Pirrone et al., 2018). In contrast to that, here we found no effects in adults with ASD. In face-related sensory processing, Powell et al. (2019) reported reduced decision boundary and reduced drift rate in adolescents with ASD. It is difficult to reconcile these findings within the general framework of sensory processing. It is possible that these findings are task-, stimuli- and/or sample-specific (see Introduction section of this chapter for more detailed discussion). Difficulties in finding robust task effects in ASD has long been one of the main hurdles in ASD research (see Robertson and Baron-Cohen, 2017, for a recent review). These problems could be largely overcome by having larger ASD samples that are stratified based on individual symptom profiles and are tested on a battery of different tasks and different stimuli within a single study.

Our finding of increased decision boundary in schizophrenia is in agreement with Heathcote et al. (2015), who also found decision boundary to be increased in schizophrenia in left-vs-right motion discrimination task. Furthermore, just like Heathcote et al. (2015) we interpreted this result as a compensatory strategy in response to poorer sensory sensitivity (in our sample poorer sensory sensitivity was indicated by SCZ converging to higher stimulus contrast levels; see **Chapter 3**). Another study (Moustafa et al., 2015), using a probabilistic reward and punishment learning task, also found SCZ to exhibit larger decision boundaries. In addition, Moustafa et al. (2015) found slower drift rates on punishment trials and longer non-decision times. This is in agreement with another recent

study using a processing-speed digit-symbol coding task (Mathias et al., 2017), where patients with schizophrenia also showed decreased drift rates and increased non-decision times. It is important to keep in mind, however, that most studies mentioned above, as well as our own study, tested chronically ill patients that had been medicated for a long time with antipsychotics (mainly dopamine antagonists) and this might affect the observed results (e.g., see Hutton et al., 2002). For example, increasing dopamine signaling pharmacologically has been shown to selectively and positively affect drift rates in the left-vs-right motion discrimination task in healthy controls (Beste et al., 2018). Furthermore, in Parkinson's disease, which is underlied by a dopaminergic hypofunction (Iversen and Iversen, 2007), a recent study has reported slower drift rates in Attention Network Task (O'Callaghan et al., 2017). This emphasizes the importance of stratifying patients based on their disease stage and conducting longitudinal studies to account for the changes that happen throughout the course of illness and in response to treatment.

7

General discussion

In this thesis we investigated possible computational basis of schizophrenia (SCZ) and autism spectrum disorder (ASD). In particular, we experimentally tested the hypothesis that SCZ and ASD are associated with impairments in Bayesian inference. To this end, we studied schizotypy and autistic traits in the general population as well as patient groups with SCZ and ASD diagnosis in a visual motion estimation task that induces perceptual priors (Chalk et al., 2010). We performed extensive computational modelling analysis to account for task behavior and to characterise individual and group differences. In this chapter we will summarize our main findings in the context of other recent work, discuss limitations and will provide some ideas for future research.

7.1 Operationalization of priors and underlying neural mechanisms

Across autistic and schizotypy traits in the general population (**Chapter 2**) as well as SCZ (**Chapter 3**) and ASD groups (**Chapter 4**) we found intact ability to rapidly acquire perceptual priors about motion direction of visual stimuli and we observed no systematic differences in the acquired priors across all these traits

and groups. Together with many other negative findings reported in the literature (summarised in **Table 1.2** and **Table 1.1 in Chapter 1**), these results speak against the idea of a general impairment in prior acquisition and/or chronically weaker priors in SCZ and ASD.

One major contribution of this thesis is the idea that operationalization of priors should not rely on biases alone, as that confounds prior and likelihood effects. To disentangle prior and likelihood effects one needs to use task designs that allow for estimation of perceptual variability. Furthermore, to quantify prior and likelihood effects and to control for other confounds such as motor noise, it is important to fit generative models to the data. In **Chapter 2**, this led us to the finding of increased sensory precision along autistic traits in the general population, which was expressed in reduced biases and reduced response variability in motion estimation task. Interestingly, in **Chapter 3** we found that while patients with SCZ showed no differences in estimation performance - and thus no differences in motion direction priors or sensory likelihoods - they exhibited reduced number of prior-based hallucinations in the detection task. Within the Bayesian framework this could be interpreted as a result of weaker priors for stimulus presence. However, in **Chapter 6** we show that these effects can be explained by increased caution in SCZ patients (discussed in more detail in the last section of this chapter). The latter result only re-emphasizes the importance of fitting generative models in order to explicitly show that the observed effects are indeed driven by priors.

Most experimental studies operationalize the prior by measuring only a single variable (such as response bias or change in reaction time) induced by manipulation of context and do not provide generative models that could account for the observed effects (see **Table 1.2** and **Table 1.1** in **Chapter 1**). While these

effects are often reported under an umbrella term as 'prior effects', the actual mechanisms producing these effects are not always clear. For example, a few studies have reported reduced adaptation aftereffects in ASD as supporting the hypothesis of weaker influence of perceptual priors (Turi et al., 2015, 2016; Lawson et al., 2018). However, within the Bayesian framework, adaptation aftereffects (i.e., repulsive biases after a prolonged exposure to a stimulus) are more consistent with changes in sensory likelihood representations rather than in priors (Stocker and Simoncelli, 2006b; Clifford et al., 2007). Some other studies have investigated prior effects in SCZ and ASD using ambiguous (Mooney) images (Teufel et al., 2015; Davies et al., 2018; Van de Cruys et al., 2018; Król and Król, 2019), where disambiguation of two-tone images is facilitated by a single presentation of the natural source image. While these effects are often explained within the Bayesian framework with the natural source image acting as a prior, it is not clear what information such prior encodes and how it could be represented in a computational model. Other studies have investigated the effects of naturally existing long-term priors that are thought to underlie various perceptual illusions (Powell et al., 2016; Pell et al., 2016; Croydon et al., 2017; Laeng et al., 2018; Grzeczkowski et al., 2018; Utzerath et al., 2019; Kaliuzhna et al., 2019; Sandhu et al., 2020). While detailed Bayesian models have been put forward for some of the illusions (e.g., Weiss et al., 2002; Shams et al., 2005; Dimova and Denham, 2010; Brown and Friston, 2012), the exact mechanisms underlying most classical illusions remain unclear. Furthermore, multiple studies have found illusory magnitude to be uncorrelated across most illusions (Yang et al., 2013; Chouinard et al., 2016; Grzeczkowski et al., 2017; Cretenoud et al., 2019), which suggests that different illusions are underlied by different mechanisms and their strength cannot be explained by a common factor such as 'reliance on priors'.

Future research would benefit from a more rigorous operationalization of pri-

ors and from more explicit links to the underlying mechanisms that are thought to implement such priors. For example, one mechanism that can implement context-dependent modulation is *divisive normalization*, where the neural response of a given neuron is normalized by the activity of the neuronal population in which the neuron is embedded (e.g., Carandini and Heeger, 2012; Louie et al., 2013). Following recent Bayesian accounts of ASD, it has been proposed that weaker priors could manifest as reduced divisive normalization (Rosenberg et al., 2015). However, recent experimental studies have reported intact divisive normalization in ASD in low-level vision (Van de Cruys et al., 2018; Sandhu et al., 2020), suggesting that the potential impairments might arise at higher levels of processing. While divisive normalization describes a specific computation that is performed across the brain, it can nevertheless be implemented by different neurobiological mechanisms at different levels of processing (Carandini and Heeger, 2012). More distal contextual modulation, on the other hand, has been associated with different neuromodulatory systems (Friston, 2008; Iglesias et al., 2017) and can thus provide a more direct link to the neurobiology. For instance, in a relatively recent study, Lawson et al. (2017) showed a relationship between increased noradrenergic response (based on pupillometry measures), increased (model-derived) surprise about environmental volatility and reduced behavioral surprise in ASD - demonstrating how differences in behavior, model-derived quantities and neurobiology can be interrelated in a single experiment. Theoretical work has also argued for the importance of oxytocin in explaining context-dependent modulation deficits in ASD (Quattrocki and Friston, 2014); while some evidence also exists for NDMA (Lee et al., 2015) and dopamine (Pavăl, 2017) dysfunction, but no experimental work has investigated these deficits within the Bayesian framework. In SCZ, while abnormalities in dopaminergic, NMDA receptor and GABAergic systems are well documented and play an important role in supporting Bayesian accounts of the disorder (reviewed in Valton et al., 2017), the involvement of these

systems in the tasks used to test Bayesian accounts is rarely made explicit (see studies in **Table 1.1** in **Chapter 1**), with the exception of a few studies (Jardri et al., 2017; Cassidy et al., 2018; Adams et al., 2018).

Careful consideration of the underlying neural mechanisms was also absent in our choice of the Moving Dots task. Reflecting on it *post hoc*, the most important feature of our task was the low-contrast motion stimulus. In the brain, such stimulus would primarily be transmitted via magnocellular pathway projecting to the dorsal visual stream (Merigan et al., 1991). Interestingly, magnocellular pathway has been implicated in SCZ (e.g., Butler et al., 2007; Kim et al., 2006). Magnocellular deficits in SCZ are reflected in patients showing higher contrast, motion speed and direction thresholds. This could be explained by NMDA hypofunction in the magnocellular pathway, where NMDA plays an important role in gain control, amplifying weak signals resulting from low-contrast stimulus and amplifying lateral inhibition effects (e.g., Butler et al., 2008). This would also be in line with our finding of increased contrast thresholds in SCZ group (see **Chapter 3**). Some evidence exists that magnocellular pathway abnormalities feature in ASD too, with some studies reporting impaired magnocellular pathway function (e.g., Sutherland and Crewther, 2010) and some reporting enhanced function (McCleery et al., 2007). Our findings of increased sensory precision along AQ might reflect enhanced magnocellular functioning expressed as a better discrimination of motion direction, possibly via increased lateral inhibition effects. Considering the possibility that our findings are specific to magnocellular pathway abnormalities, one has to be careful generalizing them to other contexts and tasks that can be explained within the Bayesian inference framework, but which target different neural circuits.

One powerful approach for identifying mechanisms that are impaired without

necessarily having a detailed a priori understanding of them is to use a battery of tasks within a single study and investigate their inter-correlations. Applied in the context of ASD and SCZ, this might help avoid over-generalization of findings and would directly test the relationship between effects observed in different tasks. For example, reduced susceptibility to illusions is often cited as a feature of both SCZ and ASD, by selectively referring to studies with positive findings. However, studies that have systematically investigated response to illusions using batteries of different illusions, have found that in both SCZ and ASD reduced susceptibility to illusions is more of an exception than a rule. For instance, Grzeczkowski et al. (2018) used a battery of 10 classical illusions and found SCZ patients to show reduced susceptibility to only one: Simultaneous Contrast illusion. Similarly, in Yang et al. (2013) study, out of 5 tested illusions only Simultaneous Contrast illusion was found to be reduced in SCZ, while in Tibber et al. (2013) study, out of 4 tested illusions, only illusions defined by contrast and size were found to be weaker in SCZ. In ASD literature, Chouinard et al. (2016) tested 13 illusions in the general population and found only two, Square-diamond and Shepard's tabletops illusions to be reduced as a function of autistic traits. In clinical ASD, Ropar and Mitchell (1999) tested 4 illusions and found no reduced susceptibility. In another study, using a battery of visuospatial tasks in which ASD show enhanced performance, Ropar and Mitchell (2001) also found it to be not predictive of susceptibility to these visual illusions. While providing mostly negative results so far, such studies are important for directly testing whether task effects that share the same theoretical explanation are actually underlied by the same mechanisms.

7.2 Traits, diagnostic categories and symptom profiles

One important inconsistency in our findings concerns increased sensory precision along autistic traits in the general population (**Chapter 2**), while in people with ASD diagnosis we found sensory precision to be not different from that of controls (**Chapter 4**). There are several potential explanations that can be entertained to explain this inconsistency. For example, it might relate to the heterogeneity of the ASD population (e.g., see Masi et al., 2017) combined with the fact that our sample was rather small (N=17) and consisted only of high-functioning adults. Another possible explanation might concern the relationship between autistic traits in the general population and clinical ASD. While there is plenty of research supporting the continuity between sub-clinical autistic traits - as measured by Autism Quotient (AQ) questionnaire - and clinical ASD diagnosis (e.g., Baron-Cohen et al., 2001; Horder et al., 2014; Mayer, 2017), there are also multiple studies reporting divergent findings (e.g., see Ewbank et al., 2014, 2017; Lawson et al., 2018, for recent studies in the context of Bayesian framework), with some scholars suggesting that the traits captured by AQ might not always be underlied by the ASD endophenotype, leading to the inconsistencies in findings (Gregory and Plaisted-Grant, 2016). The latter might also explain why we did not replicate the finding of increased sensory precision along AQ in our control group in **Chapter 4**. In fact, we found substantial evidence against the positive relationship between AQ and sensory precision ($BF_{01} = 4.4$ for controls, and $BF_{01} = 9.6$ pooling data across controls and ASD). This suggests that more comprehensive assessment of autistic traits might be needed in future studies.

Our findings of intact prior acquisition and reduced hallucinations in patients with SCZ (**Chapter 3**) require similar considerations related to the symptom

profiles of the tested patients. Our sample consisted of chronic patients that had been medicated for a long time and did not have strong positive symptoms at the time of testing. Therefore, it is unclear if our results would generalize to more acute SCZ. Most other studies reporting positive findings reported effects associated with either hallucinations or delusions (see **Table 1.1** in **Chapter 1**). Related to that, the observed effects might also vary based on SCZ phase, with prodromal and recovery phases likely to be linked with weaker priors, while acute phase (when the positive symptoms are strongest) is more likely to be linked with stronger priors (e.g., Adams et al., 2013; Diaconescu et al., 2019). Reduced prior-based hallucinations found in our sample (in recovery phase) are in line with this view. Studying larger samples of patients, stratified based on the above mentioned factors, might further clarify how the inference mechanism changes during different phases of illness and in response to treatment.

7.3 Exploratory models

In the second part of this thesis (**Chapters 5** and **6**), we performed exploratory modelling analysis to account for additional effects present in our data. In **Chapter 5** we found that in addition to the attractive Bayesian biases induced via stimulus statistics, participants exhibited repulsive biases. One source of repulsive biases was the cardinal directions. While not well understood, repulsion from cardinals has been documented in the literature (Rauber and Treue, 1998), and is thought to be an instance of a more general phenomenon referred to as reference repulsion (for a review, see Kuang, 2019) - where repulsive biases are observed with respect to a reference direction, which is either cued, or is implicit (as in the case of cardinal directions). In addition to repulsion from the cardinals, we found evidence for repulsion from the central reference angle. In our task, the central reference angle is selected randomly (between 0° and 360°) for each participant;

and the presented stimulus (between -64° and 64°) is organized around this angle. This means that this central reference angle would need to be inferred from the statistics of the stimulus. We are not aware of any other work showing that reference repulsion could be induced in such an indirect way. Furthermore, the support for this effect in our analysis came from fixed-effects model comparison, while random-effects BMS was inconclusive. Therefore, replication studies would be needed to verify the existence of such an effect.

Interestingly, however, we found repulsion from the central reference angle to be reduced along schizotypy traits in the general population as well as in SCZ and ASD groups. While an exciting finding in its consistency across traits and clinical groups, its significance for the Bayesian accounts of these disorders is hard to gauge due to there not being an established Bayesian explanation of these effects. However, as we reasoned above, the central reference angle could only be inferred from accurate estimation of stimulus statistics and thus could relate to differences in prior acquisition, even though no differences in the priors themselves were found.

In **Chapter 6**, to better understand the effects of reaction time (and exposure to stimulus duration) on the observed Bayesian effects, we devised a Continuous Choice Drift Diffusion Model (CDM). We found that increased sensory precision found along autistic traits in the Bayesian model was underlied by faster accumulation of sensory evidence, while slower reaction times and less hallucinations in SCZ were a result of increased decision boundary. However, no single model parameter was found to be responsible for faster reaction times in ASD group, and was likely a collective effect of decision boundary, starting point distance and non-decision time, as none of these parameters had a strong support for the null hypothesis ($BF_{01} = 1.4$, $BF_{01} = 1.8$ and $BF_{01} = 2.7$, respectively).

More generally, our within-trial dynamics analysis demonstrates the importance of accounting for exposure to stimulus time when studying Bayesian biases. Our model simulations showed that all parameters that affect bias, (decision boundary, drift rate and starting point distance) also affect reaction times. Thus, if reaction times or exposure to stimulus duration are not accounted for, it might lead to confounded results. In different tasks used to study ASD and SCZ, exposure to stimulus duration varies from 200ms (Lawson et al., 2018), to 13.6s (Utzerath et al., 2019), while in studies that have tested batteries of illusions in SCZ (Grzeczkowski et al., 2018; Yang et al., 2013) and in ASD (Chouinard et al., 2016; Ropar and Mitchell, 2001), unlimited time is usually given to respond on each trial. Meanwhile, it has been shown that the magnitude of visual illusions can vary as a function of exposure time (Calvert and Harris, 1988) and that reduced illusion effects along schizotypy traits can be explained by increased reaction times (Bressan and Kramer, 2013). In general, only a handful of studies have investigated within-trial dynamics in ASD and SCZ (in ASD: Pirrone et al., 2017, 2018; Karalunas et al., 2018; Powell et al., 2019; in SCZ: Heathcote et al., 2015; Moustafa et al., 2015; Mathias et al., 2017), despite the differences in processing speed being commonly reported in both SCZ (Schatz, 1998) and ASD (Robertson and Baron-Cohen, 2017). Thus, investigating within-trial dynamics within the context of Bayesian inference is a promising direction for future research in SCZ and ASD.

Finally, it is worth noting that some recent Bayesian studies of SCZ and ASD have investigated the dynamics of how sensory evidence is integrated over time, but instead of within-trial dynamics they focused on across-trial dynamics, with an emphasis on prior updating. This would include studies discussed throughout this thesis using sequential sampling tasks (Cassidy et al., 2018; Adams et al.,

2018; Baker et al., 2019), and associative learning under volatile conditions (Powers et al., 2017; Lawson et al., 2017). In our task, the priors were acquired relatively quickly (within 100-170 trials) and stayed relatively stable throughout the rest of the task (e.g., see **Supplementary Fig. 3** in **Chapter 4**), preventing the possibility of studying individual and group differences in the dynamics of prior updating in great detail.

Appendix A

Supplementary material for Chapter 5

A.1 Behavioral measures in absolute motion direction coordinates

To gain a better insight into how cardinal directions were affecting performance in our task, in addition to analysing estimation biases (presented in the main text in Chapter 5), we also analysed all other behavioral measures in absolute motion direction coordinates. While the effects on estimation bias showed clear periodicity every 90° (**Fig. A.1A**), most other measures were not significantly affected by the cardinals: variability (main motion direction effect: $F(22, 48.82) = 1.32,\ p = .205$; **Fig. A.1B**), lapse estimations ($F(22, 15.59) = 1.08,\ p = .445$; **Fig. A.1C**), reaction times ($F(22, 50.43) = 1.32,\ p = .203$; **Fig. A.1D**) and hallucinations ($F(22, 1886) = 0.96,\ p = .518$; **Fig. A.1F**). Only detection rate was significantly affected by the cardinals ($F(22, 47.13) = 6.81,\ p < .001$; **Fig. A.1E**), with the detection being better for horizontal directions (around $-90°$ and $90°$) compared to performance at vertical directions (around $0°$ and $\pm 180°$). We are unaware of any previous literature reporting better detection for horizontal than vertical

directions and it is unclear how it might relate to the observed repulsive biases, as our models did not explicitly account for the detection performance.

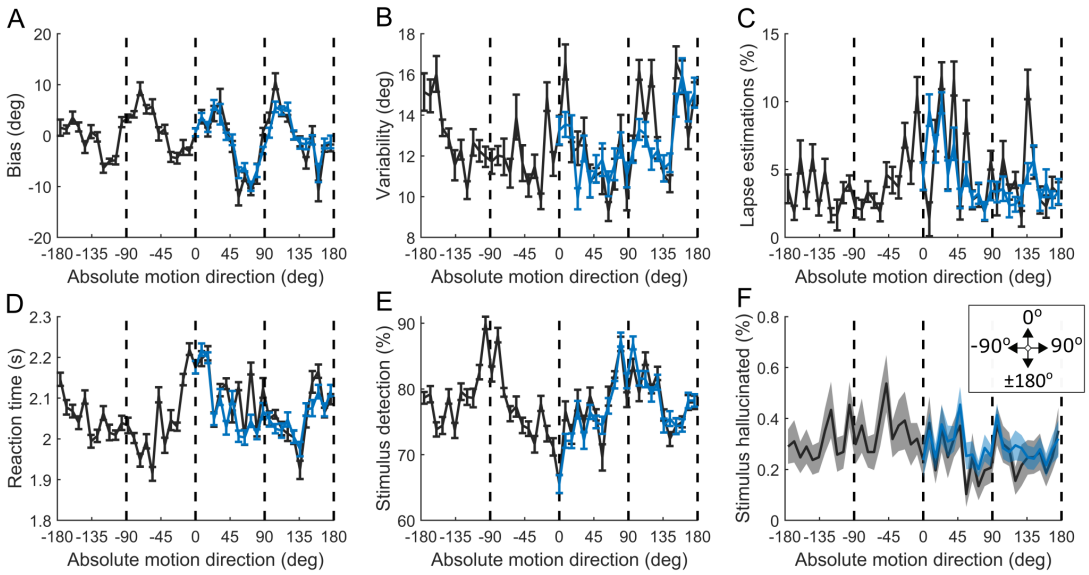


Figure A.1: Behavioral measures in absolute motion direction coordinates for the general population sample ($N = 83$) on low-contrast trials (**A-E**) and no-stimulus trials (**F**). (**A**) Estimation bias, (**B**) estimation variability, (**C**) lapse estimations, (**D**) estimation reaction time, (**E**) detection rate on low-contrast trials, (**F**) hallucination rate on no-stimulus trials. Black line represents data across the whole range of motion directions, blue line represents the same data but folded with respect to 0° (averaging points at an equal absolute distance from the 0° angle). The vertical dashed lines indicate the cardinal directions. The error bars and shaded areas represent within-subject standard error.

Most of the research on the effects of the cardinal directions in motion perception, has focused on the effects in discrimination (as opposed to detection) performance. Here, the effects of cardinals have been shown to be expressed as an advantage in discriminating cardinal vs oblique directions (*the oblique effect*) (e.g., see Gros et al., 1998; Dakin et al., 2005b; Greenwood and Edwards, 2007); while more recently, improved discrimination has also been reported for horizontal vs vertical motion directions (Pilz and Papadaki, 2019). We do not see these effects in our data (as suggested by no significant effects in estimation variability). One possible explanation for this discrepancy might lie in the distinction

between *stimulus noise* (i.e., sensory ambiguity induced by reduced coherence of motion), which was used by studies investigating the oblique effect, and *sensory noise* (i.e. sensory ambiguity induced by reduced contrast or presentation time), which was used in our studies. For example, Wei and Stocker (2015) have shown that under the assumption of efficient coding, stimulus noise would lead to attractive biases towards the cardinals, while sensory noise would lead to repulsive biases from the cardinals. In our data we do observe repulsive biases, in line with Wei and Stocker results. While these differential effects were only demonstrated for biases, we might speculate it to be relevant for explaining the lack of effects in estimation variability in our data. More generally, such results also emphasize the importance of recognizing different sources of uncertainty and considering the possibility that it might have different roles in the computations performed by the brain.

A.2 Complete model fits for repulsion analysis

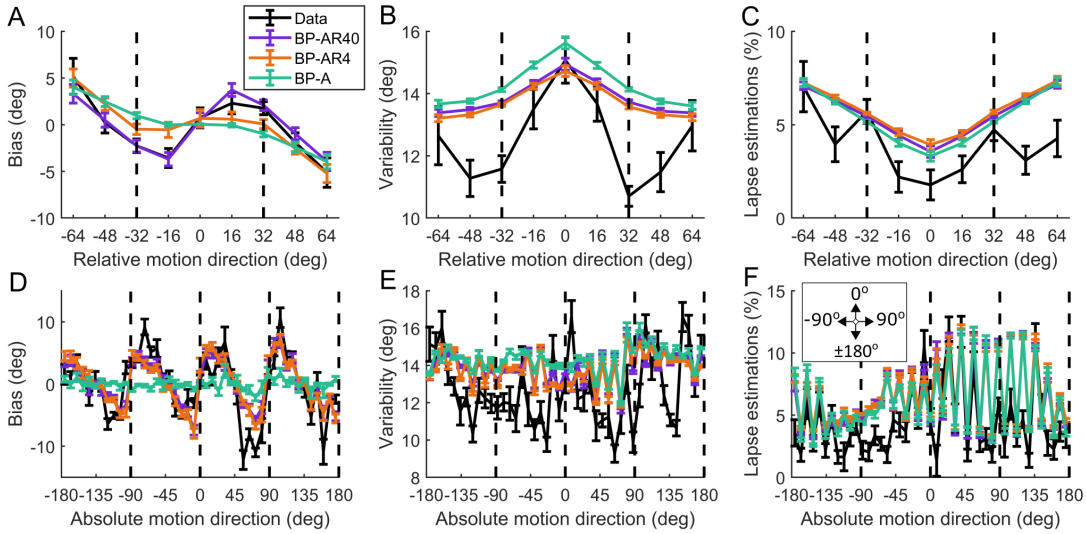


Figure A.2: BP-AR40, BP-AR4 and BP-A models' fits for all estimation performance measures in relative motion direction (**A-C**) and absolute motion direction coordinates (**D-F**). (**A,D**) Estimation bias, (**B,E**) estimation variability, (**C,F**) lapse estimations. The vertical dashed lines in **A-C** indicate the two most frequently presented directions, while in **D-F** they indicate the cardinal directions. The error bars represent within-subject standard error.

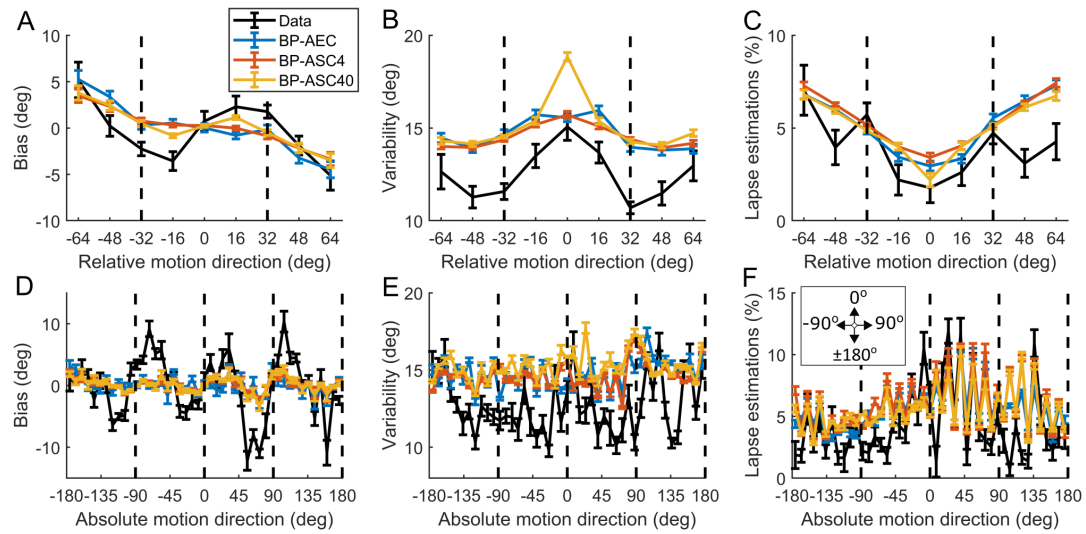


Figure A.3: BP-AEC, BP-ASC4 and BP-ASC40 models' fits for all estimation performance measures in relative motion direction (**A-C**) and absolute motion direction coordinates (**D-F**). (**A,D**) Estimation bias, (**B,E**) estimation variability, (**C,F**) lapse estimations. The vertical dashed lines in **A-C** indicate the two most frequently presented directions, while in **D-F** they indicate the cardinal directions. The error bars represent within-subject standard error.

A.3 Repulsion effects in Chalk et al. (2010) data

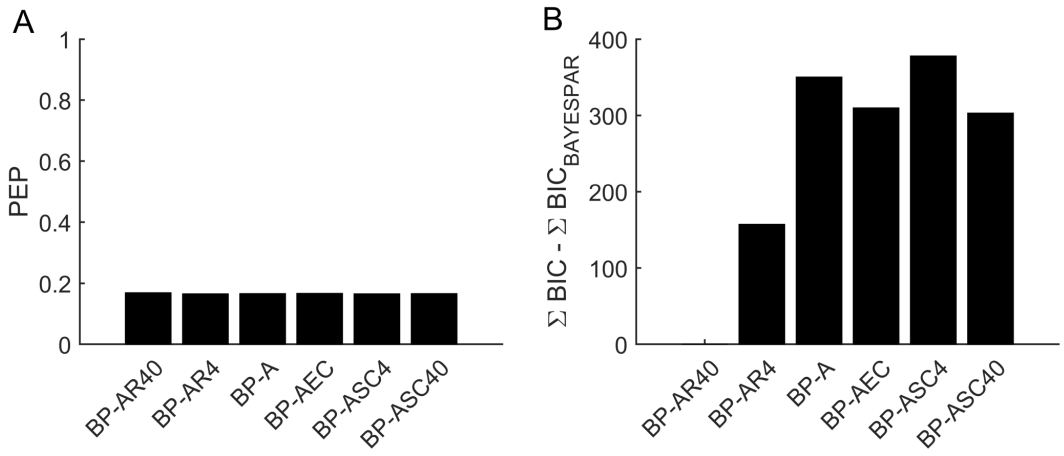


Figure A.4: Model comparison for repulsion analysis with Chalk et al. (2010) data. (**A**) Random-effects Bayesian model selection results; PEP - protected exceedance probability. (**B**) Fixed-effects model comparison. Note that the plot shows the difference between the model with the lowest BIC and the remaining models. BP-AR40 had the lowest BIC with 157 units lead.

Appendix B

Supplementary material for Chapter 6

B.1 Reaction time within-subject effects by CDM

CDM predicted reaction time effect found in behavioral data, with faster trials being associated with more bias (**Fig. B.1A**). However, the model also predicted more lapse estimations for faster trials (**Fig. B.1C**), which we did not see in behavioral data.

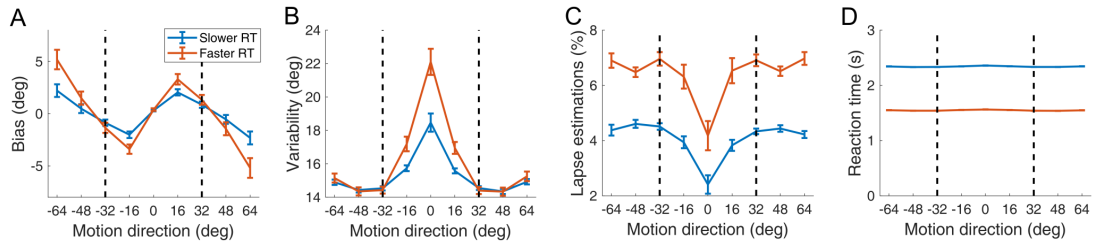


Figure B.1: Reaction time within-subject effects predicted by CDM: fast vs slow trials (35% fastest and 35% slowest trials for each individual). (**A**) Estimation bias, (**B**) estimation variability, (**C**) lapse estimations, (**D**) reaction time. The vertical dashed lines correspond to the two most frequently presented motion directions ($\pm 32^{\circ}$). Error bars represent within-subject standard error.

B.2 CDM parameter differences between fast and slow participants

Analysing CDM parameter difference between 30 fastest and 30 slowest individuals revealed that faster individuals had smaller decision boundary and more lapse estimations (**Fig. B.2**). Together, this can be interpreted as faster individuals being less cautious when performing the task.

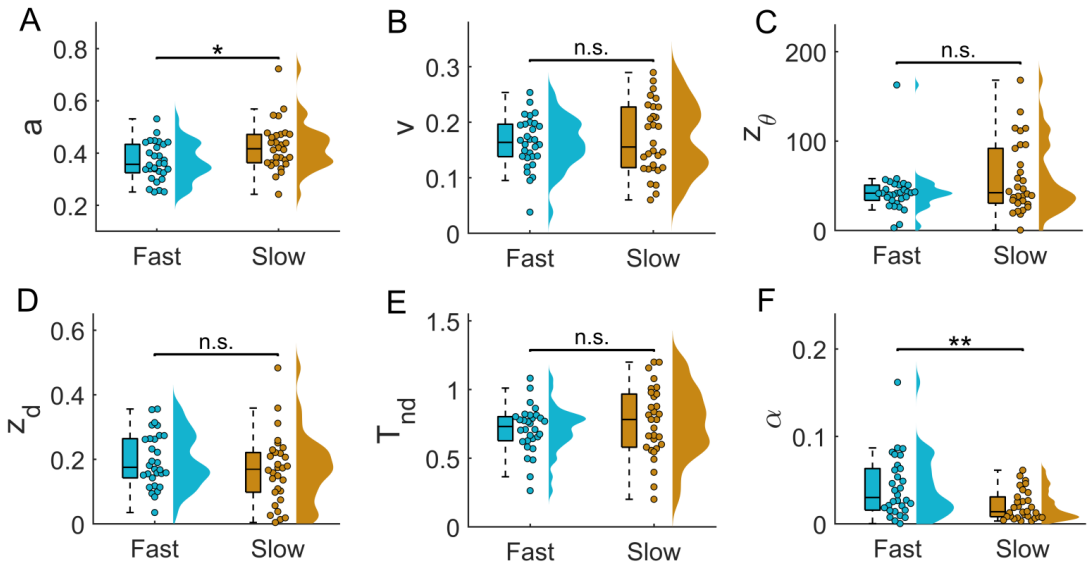


Figure B.2: CDM parameter comparison between 30 fastest and 30 slowest participants. (**A**) a - decision boundary, (**B**) v - mean drift rate, (**C**) z_θ - starting point angle, (**D**) z_d - starting point distance, (**E**) T_{nd} - non-decision time, (**F**) α - random lapse responses. Two-tailed Mann-Whitney U test results are reported with each panel. Asterisks indicate statistical significance: * - $p < .05$, ** - $p < .01$.

B.3 Within-trial dynamics in Chalk et al. (2010) data

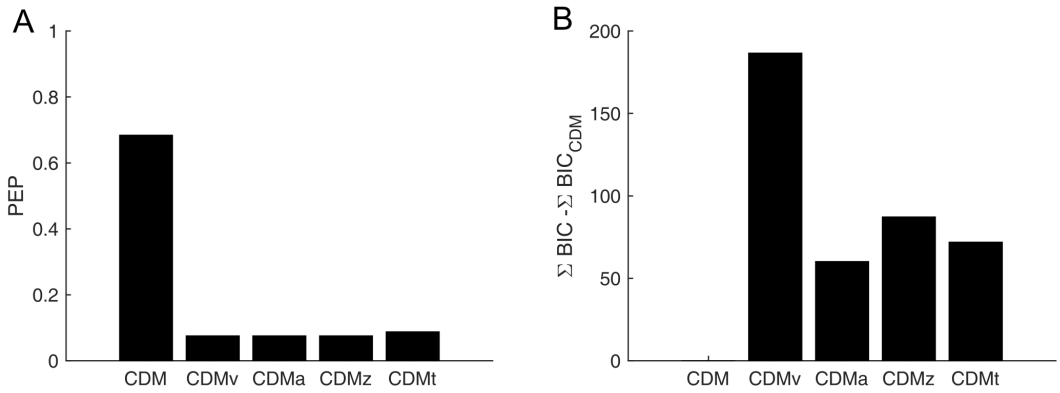


Figure B.3: Model comparison for repulsion analysis with data from Chalk et al. (2010). (**A**) Random-effects Bayesian model selection results; PEP - protected exceedance probability. (**B**) Fixed-effects model comparison. Note that the plot shows the difference between the model with the lowest BIC and the remaining models. Both methods indicate CDM to be the best model.

B.4 CDM fits to SCZ and ASD patient groups

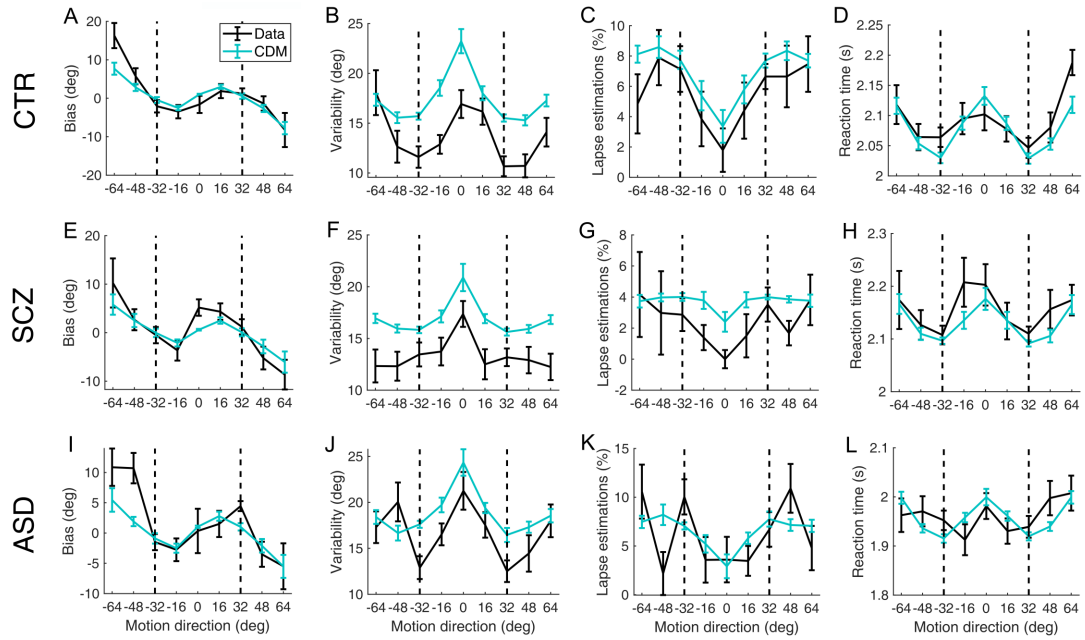


Figure B.4: CDM fits to data for controls (**A-D**), patients with schiziphrenia (**E-H**) and patients with ASD (**I-L**). (**A, E, I**) Estimation bias, (**B, F, J**) estimation variability, (**C, G, K**) lapse estimations, (**D, H, L**) estimation reaction time. The vertical dashed lines denote the two most frequently presented directions. Error bars represent within-subject standard error.

Bibliography

Acerbi, L., Vijayakumar, S., and Wolpert, D. M. (2014). On the origins of suboptimality in human probabilistic inference. *PLoS computational biology*, 10(6):e1003661.

Adams, R. A., Huys, Q. J., and Roiser, J. P. (2016). Computational psychiatry: towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery & Psychiatry*, 87(1):53–63.

Adams, R. A., Napier, G., Roiser, J. P., Mathys, C., and Gilleen, J. (2018). Attractor-like dynamics in belief updating in schizophrenia. *Journal of Neuroscience*, 38(44):9471–9485.

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., and Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in psychiatry*, 4.

Adams, W. J., Graf, E. W., and Ernst, M. O. (2004). Experience can change the'light-from-above'prior. *Nature neuroscience*, 7(10):1057–1058.

Aitchison, L. and Lengyel, M. (2017). With or without you: predictive coding and bayesian inference in the brain. *Current opinion in neurobiology*, 46:219–227.

Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, 14(3):257–262.

Alderson-Day, B., Lima, C. F., Evans, S., Krishnan, S., Shanmugalingam, P., Fernyhough, C., and Scott, S. K. (2017). Distinct processing of ambiguous speech in people with non-clinical auditory verbal hallucinations. *Brain*, 140(9):2475–2489.

Allen, P., Aleman, A., and Mcguire, P. K. (2007). Inner speech models of auditory verbal hallucinations: evidence from behavioural and neuroimaging studies. *International Review of Psychiatry*, 19(4):407–415.

Amoruso, L., Narzisi, A., Pinzino, M., Finisguerra, A., Billeci, L., Calderoni, S., Fabbro, F., Muratori, F., Volzone, A., and Urgesi, C. (2019). Contextual priors do not modulate action prediction in children with autism. *Proceedings of the Royal Society B*, 286(1908):20191319.

Andrews, T. J. and Schluppeck, D. (2000). Ambiguity in the perception of moving stimuli is resolved in favour of the cardinal axes. *Vision research*, 40(25):3485–3493.

Angela, J. Y. and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4):681–692.

APA (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Asai, T., Sugimori, E., Bando, N., and Tanno, Y. (2011). The hierarchic structure in schizotypy and the five-factor model of personality. *Psychiatry Research*, 185(1):78–83.

Association, A. P., Association, A. P., et al. (2000). Diagnostic and statistical manual of mental disorders (revised 4th ed.). *Washington, DC: American Psychiatric Association*.

Association, A. P. et al. (1987). *Diagnostic and Statistical Manual of Mental Health Disorders (DSM-III-R)*. American psychiatric association.

Austin, M.-P., Mitchell, P., and Goodwin, G. M. (2001). Cognitive deficits in depression. *The British Journal of Psychiatry*, 178(3):200–206.

Baker, S. C., Konova, A. B., Daw, N. D., and Horga, G. (2019). A distinct inferential mechanism for delusions in schizophrenia. *Brain*, 142(6):1797–1812.

Barlow, H. (1969). Pattern recognition and the responses of sensory neurons. *Annals of the New York Academy of Sciences*.

Barneveld, P. S., Pieterse, J., de Sonneville, L., van Rijn, S., Lahuis, B., van Engeland, H., and Swaab, H. (2011). Overlap of autistic and schizotypal traits in adolescents with autism spectrum disorders. *Schizophrenia research*, 126(1):231–236.

Baron-Cohen, S. (2000). Theory of mind and autism: A fifteen year review. *Understanding other minds: Perspectives from developmental cognitive neuroscience*, 2:3–20.

Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T., and Chakrabarti, B. (2009). Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1522):1377–1383.

Baron-Cohen, S., Leslie, A. M., Frith, U., et al. (1985). Does the autistic child have a "theory of mind". *Cognition*, 21(1):37–46.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (aq): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders*, 31(1):5–17.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, 76(4):695–711.

Becchio, C., Mari, M., and Castiello, U. (2010). Perception of shadows in children with autism spectrum disorders. *PloS one*, 5(5):e10582.

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., Shadlen, M. N., Latham, P. E., and Pouget, A. (2008). Probabilistic population codes for bayesian decision making. *Neuron*, 60(6):1142–1152.

Behrmann, M., Thomas, C., and Humphreys, K. (2006). Seeing it differently: visual processing in autism. *Trends in cognitive sciences*, 10(6):258–264.

Beierholm, U. R., Quartz, S. R., and Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of vision*, 9(5):23–23.

Bertone, A., Mottron, L., Jelenic, P., and Faubert, J. (2003). Motion perception in autism: a "complex" issue. *Journal of cognitive neuroscience*, 15(2):218–225.

Bertone, A., Mottron, L., Jelenic, P., and Faubert, J. (2005). Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity.

Brain, 128(10):2430–2441.

Beste, C., Adelhöfer, N., Gohil, K., Passow, S., Roessner, V., and Li, S.-C. (2018). Dopamine modulates the efficiency of sensory evidence accumulation during perceptual decision making. *International Journal of Neuropsychopharmacology*, 21(7):649–655.

Bitzer, S., Park, H., Blankenburg, F., and Kiebel, S. J. (2014). Perceptual decision making: drift-diffusion model is equivalent to a bayesian model. *Frontiers in human neuroscience*, 8:102.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, 113(4):700.

Boulter, C., Freeston, M., South, M., and Rodgers, J. (2014). Intolerance of uncertainty as a framework for understanding anxiety in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 44(6):1391.

Bowers, J. S. and Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological bulletin*, 138(3):389.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10:433–436.

Bressan, P. and Kramer, P. (2013). The relation between cognitive-perceptual schizotypal traits and the ebbinghaus size-illusion is mediated by judgment time. *Frontiers in psychology*, 4:343.

Brock, J. (2012). Alternative bayesian accounts of autistic perception: comment on pellicano and burr. *Trends in cognitive sciences*, 16(12):573–574.

Brodersen, K. H., Deserno, L., Schlagenhauf, F., Lin, Z., Penny, W. D., Buhmann, J. M., and Stephan, K. E. (2014). Dissecting psychiatric spectrum disorders by generative embedding. *NeuroImage: Clinical*, 4:98–111.

Brosnan, M., Chapman, E., and Ashwin, C. (2014). Adolescents with autism spectrum disorder show a circumspect reasoning bias rather than 'jumping-to-conclusions'. *Journal of autism and developmental disorders*, 44(3):513–520.

Brown, H. and Friston, K. J. (2012). Free-energy and illusions: the cornsweet effect. *Frontiers in psychology*, 3:43.

Buchy, L., Woodward, T. S., and Liotti, M. (2007). A cognitive bias against disconfirmatory evidence (bade) is associated with schizotypy. *Schizophrenia research*, 90(1):334–337.

Burge, J., Fowlkes, C. C., and Banks, M. S. (2010). Natural-scene statistics predict how the figure–ground cue of convexity affects human depth perception. *Journal of Neuroscience*, 30(21):7269–7280.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304.

Butler, P. D., Martinez, A., Foxe, J. J., Kim, D., Zemon, V., Silipo, G., Mahoney, J., Shpaner, M., Jalbrzikowski, M., and Javitt, D. C. (2007). Subcortical visual dysfunction in schizophrenia drives secondary cortical impairments. *Brain*, 130(2):417–430.

Butler, P. D., Silverstein, S. M., and Dakin, S. C. (2008). Visual perception and its impairment in schizophrenia. *Biological psychiatry*, 64(1):40–47.

Calvert, J. and Harris, J. (1988). Spatial frequency and duration effects on the tilt illusion and orientation acuity. *Vision Research*, 28(9):1051–1059.

Carandini, M. and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51.

Cassidy, C. M., Balsam, P. D., Weinstein, J. J., Rosengard, R. J., Slifstein, M., Daw, N. D., Abi-Dargham, A., and Horga, G. (2018). A perceptual inference mechanism for hallucinations linked to striatal dopamine. *Current Biology*, 28(4):503–514.

Chalk, M., Seitz, A. R., and Seriès, P. (2010). Rapidly learned stimulus expectations alter perception of motion. *Journal of Vision*, 10(8):2–2.

Chambon, V., Farrer, C., Pacherie, E., Jacquet, P. O., Leboyer, M., and Zalla, T. (2017). Reduced sensitivity to social priors during action prediction in adults with autism spectrum disorders. *Cognition*, 160:17–26.

Chen, Y. (2011). Abnormal visual motion processing in schizophrenia: a review of research progress. *Schizophrenia bulletin*, 37(4):709–715.

Chisholm, K., Lin, A., Abu-Akel, A., and Wood, S. J. (2015). The association between autism and schizophrenia spectrum disorders: a review of eight alternate models of co-occurrence. *Neuroscience & Biobehavioral Reviews*, 55:173–183.

Chouinard, P. A., Unwin, K. L., Landry, O., and Sperandio, I. (2016). Susceptibility to optical illusions varies as a function of the autism-spectrum quotient but not in ways predicted by local–global biases. *Journal of autism and developmental disorders*, 46(6):2224–2239.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204.

Clifford, C. W., Webster, M. A., Stanley, G. B., Stocker, A. A., Kohn, A., Sharpee, T. O., and Schwartz, O. (2007). Visual adaptation: Neural, psychological and computational aspects. *Vision research*, 47(25):3125–3131.

Cohen Hoffing, R., Karvelis, P., Rupprechter, S., Seriès, P., and Seitz, A. R. (2018). The influence of feedback on task-switching performance: A drift diffusion modeling account. *Frontiers in integrative neuroscience*, 12:1.

Colombo, M. and Seriès, P. (2012). Bayes in the brain—on bayesian modelling in neuroscience. *The British journal for the philosophy of science*, 63(3):697–723.

Conant, R. C. and Ross Ashby, W. (1970). Every good regulator of a system must be a model of that system. *International journal of systems science*, 1(2):89–97.

Constantino, J. N. and Todd, R. D. (2003). Autistic traits in the general population: a twin study. *Archives of general psychiatry*, 60(5):524–530.

Corlett, P., Frith, C. D., and Fletcher, P. (2009). From drugs to deprivation: a bayesian framework for understanding models of psychosis. *Psychopharmacology*, 206(4):515–530.

Corlett, P. R. and Fletcher, P. C. (2014). Computational psychiatry: a rosetta stone linking the brain to mental illness. *The Lancet Psychiatry*, 1(5):399–402.

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., and Powers III, A. R. (2019). Hallucinations and strong priors. *Trends in cognitive sciences*, 23(2):114–127.

Cretenoud, A. F., Karimpur, H., Grzeczkowski, L., Francis, G., Hamburger, K., and Herzog, M. H. (2019). Factors underlying visual illusions are illusion-specific but not feature-specific. *Journal of Vision*, 19(14):12–12.

Croydon, A., Karaminis, T., Neil, L., Burr, D., and Pellicano, E. (2017). The light-from-above prior is intact in autistic children. *Journal of Experimental Child Psychology*, 161:113–125.

Cuthbert, B. N. (2014). The rdoc framework: facilitating transition from icd/dsm to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, 13(1):28–35.

Dakin, S., Carlin, P., and Hemsley, D. (2005a). Weak suppression of visual context in chronic schizophrenia. *Current Biology*, 15(20):R822–R824.

Dakin, S. and Frith, U. (2005). Vagaries of visual perception in autism. *Neuron*, 48(3):497–507.

Dakin, S. C., Mareschal, I., and Bex, P. J. (2005b). An oblique effect for local motion: Psychophysics and natural movie statistics. *Journal of Vision*, 5(10):9–9.

Dalal, P. K. and Sivakumar, T. (2009). Moving towards icd-11 and dsm-v: Concept and evolution of psychiatric classification. *Indian Journal of Psychiatry*, 51(4):310.

Daunizeau, J., Adam, V., and Rigoux, L. (2014). Vba: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. *PLoS Computational Biology*, 10(1):e1003441.

Davies, D. J., Teufel, C., and Fletcher, P. C. (2018). Anomalous perceptions and beliefs are associated with shifts toward different types of prior knowledge in perceptual inference. *Schizophrenia bulletin*, 44(6):1245–1253.

Dayan, P., Hinton, G. E., Neal, R. M., and Zemel, R. S. (1995). The helmholtz machine. *Neural computation*, 7(5):889–904.

De Koning, M., Bloemen, O., Van Amelsvoort, T., Becker, H., Nieman, D., Van Der Gaag, M., and Linszen, D. (2009). Early intervention in patients at ultra high risk of psychosis: benefits and risks. *Acta psychiatrica scandinavica*, 119(6):426–442.

Diaconescu, A. O., Hauke, D. J., and Borgwardt, S. (2019). Models of persecutory delusions: a mechanistic insight into the early stages of psychosis. *Molecular psychiatry*, 24(9):1258–1267.

Diaconescu, A. O., Mathys, C., Weber, L. A., Kasper, L., Mauer, J., and Stephan, K. E. (2017). Hierarchical prediction errors in midbrain and septum during social learning. *Social cognitive and affective neuroscience*, 12(4):618–634.

Dickinson, A., Jones, M., and Milne, E. (2014). Oblique orientation discrimination thresholds are superior in those with a high level of autistic traits. *Journal of Autism and Developmental Disorders*, 44(11):2844.

Dima, D., Roiser, J. P., Dietrich, D. E., Bonnemann, C., Lanfermann, H., Emrich, H. M., and Dillo, W. (2009). Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. *Neuroimage*, 46(4):1180–1186.

Dimova, K. and Denham, M. (2009). A neurally plausible model of the dynamics of motion integration in smooth eye pursuit based on recursive bayesian estimation. *Biological cybernetics*, 100(3):185.

Dimova, K. D. and Denham, M. J. (2010). A model of plaid motion perception based on recursive bayesian integration of the 1-d and 2-d motions of plaid features. *Vision research*, 50(6):585–597.

Doya, K. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT press.

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

Ewbank, M. P., Pell, P. J., Powell, T. E., von dem Hagen, E. A. H., Baron-Cohen, S., and Calder, A. J. (2017). Repetition Suppression and Memory for Faces is Reduced in Adults with Autism Spectrum Conditions. *Cerebral Cortex*, 27(1):92–103.

Ewbank, M. P., Rhodes, G., von dem Hagen, E. A. H., Powell, T. E., Bright, N., Stoyanova, R. S., Baron-Cohen, S., and Calder, A. J. (2014). Repetition suppression in ventral visual cortex is diminished as a function of increasing autistic traits. *Cerebral Cortex*, 25(10):3381–3393.

Ewbank, M. P., von dem Hagen, E. A., Powell, T. E., Henson, R. N., and Calder, A. J. (2015). The effect of perceptual expectation on repetition suppression to faces is not modulated by variation in autistic traits. *Cortex*, 80:51–60.

Fard, P. R., Park, H., Warkentin, A., Kiebel, S. J., and Bitzer, S. (2017). A bayesian reformulation of the extended drift-diffusion model in perceptual decision making. *Frontiers in computational neuroscience*, 11:29.

Feldman, J. (2013). Tuning your priors to the world. *Topics in cognitive science*, 5(1):13–34.

Fine, C., Gardner, M., Craigie, J., and Gold, I. (2007). Hopping, skipping or jumping to conclusions? clarifying the role of the jtc bias in delusions. *Cognitive neuropsychiatry*, 12(1):46–77.

Fiorillo, C. D., Tobler, P. N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, 299(5614):1898–1902.

Fiser, J., Berkes, P., Orbán, G., and Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in cognitive sciences*, 14(3):119–130.

Fletcher, P. C. and Frith, C. D. (2009). Perceiving is believing: a bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10(1):48–58.

Forstmann, B. U., Ratcliff, R., and Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual review of psychology*, 67:641–666.

Fournier, K. A., Hass, C. J., Naik, S. K., Lodha, N., and Cauraugh, J. H. (2010). Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. *Journal of autism and developmental disorders*, 40(10):1227–1240.

Franklin, A., Sowden, P., Notman, L., Gonzalez-Dixon, M., West, D., Alexander, I., Loveday, S., and White, A. (2010). Reduced chromatic discrimination in children with autism spectrum disorders. *Developmental science*, 13(1):188–200.

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, 16(9):1325–1352.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1456):815–836.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput Biol*, 4(11):e1000211.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.

Friston, K., Brown, H. R., Siemerkus, J., and Stephan, K. E. (2016). The dysconnection hypothesis (2016). *Schizophrenia research*, 176(2-3):83–94.

Friston, K. J., Lawson, R., and Frith, C. D. (2013). On hyperpriors and hypopriors: comment on pellicano and burr. *Trends in cognitive sciences*, 17(1):1.

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., Dolan, R. J., Moran, R., Stephan, K. E., and Bestmann, S. (2012). Dopamine, affordance and active inference. *PLoS computational biology*, 8(1).

Friston, K. J., Stephan, K. E., Montague, R., and Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry*, 1(2):148–158.

Frith, C. and Friston, K. J. (2013). False perceptions and false beliefs: understanding schizophrenia. *Neurosci Hum Pers New Perspect Hum Act*, pages 1–15.

Frith, U. (1989). *Autism: Explaining the enigma (2nd edition 2003)*. Blackwell Publishing.

Fritsche, M. and de Lange, F. P. (2019). Reference repulsion is not a perceptual illusion. *Cognition*, 184:107–118.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and individual differences*, 7(3):385–400.

Garcıa-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision research*, 38(12):1861–1881.

Gekas, N., Chalk, M., Seitz, A. R., and Seriès, P. (2013). Complexity and specificity of experimentally induced expectations in motion perception. *J Vis*, 13:1–18.

Girshick, A. R., Landy, M. S., and Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature neuroscience*, 14(7):926–932.

Gomot, M. and Wicker, B. (2012). A challenging, unpredictable world for people with autism spectrum disorder. *International Journal of Psychophysiology*, 83(2):240–247.

Greenwood, J. A. and Edwards, M. (2007). An oblique effect for transparent-motion detection caused by variation in global-motion direction-tuning bandwidths. *Vision Research*, 47(11):1411–1423.

Gregory, B. and Plaisted-Grant, K. (2016). The autism-spectrum quotient and visual search: Shallow and deep autistic endophenotypes. *Journal of autism and developmental disorders*, 46(5):1503–1512.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 290(1038):181–197.

Gros, B. L., Blake, R., and Hiris, E. (1998). Anisotropies in visual motion perception: a fresh look. *JOSA A*, 15(8):2003–2011.

Grzeczkowski, L., Clarke, A. M., Francis, G., Mast, F. W., and Herzog, M. H. (2017). About individual differences in vision. *Vision research*, 141:282–292.

Grzeczkowski, L., Roinishvili, M., Chkonia, E., Brand, A., Mast, F. W., Herzog, M. H., and Shaqiri, A. (2018). Is the perception of illusions abnormal in schizophrenia? *Psychiatry research*, 270:929–939.

Haddad, P. M. and Sharma, S. G. (2007). Adverse effects of atypical antipsychotics. *CNS drugs*, 21(11):911–936.

Hanes, D. P. and Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, 274(5286):427–430.

Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., and Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *Journal of Neuroscience*, 31(17):6339–6352.

Happé, F. and Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, 36(1):5–25.

Happé, F., Ronald, A., and Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature neuroscience*, 9(10):1218–1220.

Happé, F. G. (1996). Studying weak central coherence at low levels: children with autism do not succumb to visual illusions. a research note. *Journal of Child Psychology and Psychiatry*, 37(7):873–877.

Harris, H., Israeli, D., Minshew, N., Bonneh, Y., Heeger, D. J., Behrmann, M., and Sagi, D. (2015). Perceptual learning in autism: over-specificity and possible remedies. *Nature neuroscience*, 18(11):1574.

Heathcote, A., Suraev, A., Curley, S., Gong, Q., Love, J., and Michie, P. T. (2015). Decision processes and the slowing of simple choices in schizophrenia. *Journal of Abnormal Psychology*, 124(4):961.

Heinz, A., Murray, G. K., Schlagenhauf, F., Sterzer, P., Grace, A. A., and Waltz, J. A. (2019). Towards a unifying cognitive, neurophysiological, and computational neuroscience account of schizophrenia. *Schizophrenia Bulletin*, 45(5):1092–1100.

Hemsley, D. R. and Garety, P. A. (1986). The formation of maintenance of delusions: a bayesian analysis. *The British Journal of Psychiatry*, 149(1):51–56.

Horder, J., Wilson, C. E., Mendez, M. A., and Murphy, D. G. (2014). Autistic traits and abnormal sensory experiences in adults. *Journal of autism and developmental disorders*, 44(6):1461–1469.

Howe, C. Q. and Purves, D. (2002). Range image statistics can explain the anomalous perception of length. *Proceedings of the National Academy of Sciences*, 99(20):13184–13188.

Howe, C. Q. and Purves, D. (2004). Size contrast and assimilation explained by the statistics of natural scene geometry. *Journal of Cognitive Neuroscience*, 16(1):90–102.

Howe, C. Q., Yang, Z., and Purves, D. (2005). The poggendorff illusion explained by natural scene geometry. *Proceedings of the National Academy of Sciences*, 102(21):7707–7712.

Howes, O. D. and Kapur, S. (2009). The dopamine hypothesis of schizophrenia: version

iii—the final common pathway. *Schizophrenia bulletin*, 35(3):549–562.

Huang, Y. and Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5):580–593.

Huq, S., Garety, P. A., and Hemsley, D. R. (1988). Probabilistic judgements in deluded and non-deluded subjects. *The Quarterly Journal of Experimental Psychology Section A*, 40(4):801–812.

Hurst, R. M., Nelson-Gray, R. O., Mitchell, J. T., and Kwapil, T. R. (2007). The relationship of asperger's characteristics and schizotypal personality traits in a non-clinical adult sample. *Journal of Autism and Developmental Disorders*, 37(9):1711–1720.

Hutton, S., Murphy, F., Joyce, E., Rogers, R., Cuthbert, I., Barnes, T., McKenna, P., Sahakian, B., and Robbins, T. (2002). Decision making deficits in patients with first-episode and chronic schizophrenia. *Schizophrenia research*, 55(3):249–257.

Huys, Q. J., Maia, T. V., and Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature neuroscience*, 19(3):404.

Huys, Q. J., Moutoussis, M., and Williams, J. (2011). Are computational models of any use to psychiatry? *Neural Networks*, 24(6):544–551.

Hyman, S. E. (2007). Can neuroscience be integrated into the dsm-v? *Nature Reviews Neuroscience*, 8(9):725–732.

Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., den Ouden, H. E., and Stephan, K. E. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron*, 80(2):519–530.

Iglesias, S., Tomiello, S., Schneebeli, M., and Stephan, K. E. (2017). Models of neuromodulation for computational psychiatry. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(3):e1420.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., and Wang, P. (2010). Research domain criteria (rdoc): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*.

Iversen, S. D. and Iversen, L. L. (2007). Dopamine: 50 years in perspective. *Trends in neurosciences*, 30(5):188–193.

Jacob, K. (2015). Recovery model of mental illness: A complementary approach to psychiatric care. *Indian journal of psychological medicine*, 37(2):117.

Jardri, R. and Denève, S. (2013). Circular inferences in schizophrenia. *Brain*, 136(11):3227–3241.

Jardri, R., Duverne, S., Litvinova, A. S., and Denève, S. (2017). Experimental evidence for circular inference in schizophrenia. *Nature communications*, 8:14218.

JASP Team (2019). JASP (Version 0.10.0)[Computer software].

Jazayeri, M. and Movshon, J. A. (2007). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*, 446(7138):912.

Ji, N. Y. and Findling, R. L. (2015). An update on pharmacotherapy for autism spectrum disorder in children and adolescents. *Current opinion in psychiatry*, 28(2):91–101.

Johns, L. C. and Van Os, J. (2001). The continuity of psychotic experiences in the

general population. *Clinical psychology review*, 21(8):1125–1141.

Joseph, R. M., Keehn, B., Connolly, C., Wolfe, J. M., and Horowitz, T. S. (2009). Why is visual search superior in autism spectrum disorder? *Developmental science*, 12(6):1083–1096.

Kaliuzhna, M., Stein, T., Rusch, T., Sekutowicz, M., Sterzer, P., and Seymour, K. J. (2019). No evidence for abnormal priors in early vision in schizophrenia. *Schizophrenia research*, 210:245–254.

Karalunas, S. L., Hawkey, E., Gustafsson, H., Miller, M., Langhorst, M., Cordova, M., Fair, D., and Nigg, J. T. (2018). Overlapping and distinct cognitive impairments in attention-deficit/hyperactivity and autism spectrum disorder without intellectual disability. *Journal of abnormal child psychology*, 46(8):1705–1716.

Karaminis, T., Cicchini, G. M., Neil, L., Cappagli, G., Aagten-Murphy, D., Burr, D., and Pellicano, E. (2016). Central tendency effects in time interval reproduction in autism. *Scientific reports*, 6.

Karim, A. R., Proulx, M. J., and Likova, L. T. (2016). Anticlockwise or clockwise? a dynamic perception-action-laterality model for directionality bias in visuospatial functioning. *Neuroscience & Biobehavioral Reviews*, 68:669–693.

Karvelis, P., Seitz, A. R., Lawrie, S. M., and Seriès, P. (2018). Autistic traits, but not schizotypy, predict increased weighting of sensory information in bayesian visual integration. *ELife*, 7:e34115.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, 90(430):773–795.

Kendell, R. and Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American journal of psychiatry*, 160(1):4–12.

Kersten, D., Mamassian, P., and Yuille, A. (2004). Object perception as bayesian inference. *Annu. Rev. Psychol.*, 55:271–304.

Kim, D., Wylie, G., Pasternak, R., Butler, P. D., and Javitt, D. C. (2006). Magnocellular contributions to impaired motion processing in schizophrenia. *Schizophrenia research*, 82(1):1–8.

King, D. J., Hodgekins, J., Chouinard, P. A., Chouinard, V.-A., and Sperandio, I. (2017). A review of abnormalities in the perception of visual illusions in schizophrenia. *Psychonomic bulletin & review*, 24(3):734–751.

Kleffner, D. A. and Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics*, 52(1):18–36.

Knill, D. C. (2007). Learning bayesian priors for depth perception. *Journal of Vision*, 7(8):13–13.

Knill, D. C. and Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719.

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

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS one*, 2(9):e943.

Körding, K. P. and Wolpert, D. M. (2004). Bayesian integration in sensorimotor learn-

ing. *Nature*, 427(6971):244–247.

Kovács, G., Grotheer, M., Münke, L., Kéri, S., and Nenadić, I. (2019). Significant repetition probability effects in schizophrenia. *Psychiatry Research: Neuroimaging*, 290:22–29.

Król, M. and Król, M. (2019). The world as we know it and the world as it is: Eye-movement patterns reveal decreased use of prior knowledge in individuals with autism. *Autism Research*, 12(9):1386–1398.

Kuang, S. (2019). Dissociating sensory and cognitive biases in human perceptual decision-making: A re-evaluation of evidence from reference repulsion. *Frontiers in Human Neuroscience*, 13:409.

Kupfer, D. J. and Regier, D. A. (2011). Neuroscience, clinical evidence, and the future of psychiatric classification in dsm-5. *American Journal of Psychiatry*, 168(7):672–674.

Kwapil, T. R., Barrantes-Vidal, N., and Silvia, P. J. (2008). The dimensional structure of the wisconsin schizotypy scales: Factor identification and construct validity. *Schizophrenia Bulletin*, 34(3):444–457.

Kwapil, T. R., Gross, G. M., Silvia, P. J., and Barrantes-Vidal, N. (2013). Prediction of psychopathology and functional impairment by positive and negative schizotypy in the chapmans' ten-year longitudinal study. *Journal of Abnormal Psychology*, 122(3):807.

Laeng, B., Færevaag, F. S., Tanggaard, S., and von Tetzchner, S. (2018). Pupillary responses to illusions of brightness in autism spectrum disorder. *i-Perception*, 9(3):2041669518771716.

Laidlaw, T. M., Dwivedi, P., Naito, A., and Gruzelier, J. H. (2005). Low self-directedness (tci), mood, schizotypy and hypnotic susceptibility. *Personality and Individual Differences*, 39(2):469–480.

Lam, K. S., Aman, M. G., and Arnold, L. E. (2006). Neurochemical correlates of autistic disorder: a review of the literature. *Research in developmental disabilities*, 27(3):254–289.

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

Lawson, R. P., Aylward, J., Roiser, J. P., and Rees, G. (2018). Adaptation of social and non-social cues to direction in adults with autism spectrum disorder and neurotypical adults with autistic traits. *Developmental cognitive neuroscience*, 29:108–116.

Lawson, R. P., Aylward, J., White, S., and Rees, G. (2015). A striking reduction of simple loudness adaptation in autism. *Scientific reports*, 5(1):1–7.

Lawson, R. P., Mathys, C., and Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature neuroscience*, 20(9):1293.

Lawson, R. P., Rees, G., and Friston, K. J. (2014). An aberrant precision account of autism. *Front Hum Neurosci*, 8.

Lee, E.-J., Choi, S. Y., and Kim, E. (2015). Nmda receptor dysfunction in autism spectrum disorders. *Current opinion in pharmacology*, 20:8–13.

Lee, P. S., Foss-Feig, J., Henderson, J. G., Kenworthy, L. E., Gilotty, L., Gaillard,

W. D., and Vaidya, C. J. (2007). Atypical neural substrates of embedded figures task performance in children with autism spectrum disorder. *Neuroimage*, 38(1):184–193.

Lee, T. S. and Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *JOSA A*, 20(7):1434–1448.

Lieder, I., Adam, V., Frenkel, O., Jaffe-Dax, S., Sahani, M., and Ahissar, M. (2019). Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nature neuroscience*, 22(2):256–264.

Lord, C., Cook, E. H., Leventhal, B. L., and Amaral, D. G. (2000). Autism spectrum disorders. *Neuron*, 28(2):355–363.

Louie, K., Khaw, M. W., and Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, 110(15):6139–6144.

Luu, L. and Stocker, A. A. (2018). Post-decision biases reveal a self-consistency principle in perceptual inference. *Elife*, 7:e33334.

Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature neuroscience*, 9(11):1432–1438.

Mamassian, P., Knill, D. C., and Kersten, D. (1998). The perception of cast shadows. *Trends in cognitive sciences*, 2(8):288–295.

Mamassian, P. and Landy, M. S. (2001). Interaction of visual prior constraints. *Vision research*, 41(20):2653–2668.

Manjaly, Z. M., Bruning, N., Neufang, S., Stephan, K. E., Brieber, S., Marshall, J. C., Kamp-Becker, I., Remschmidt, H., Herpertz-Dahlmann, B., Konrad, K., et al. (2007). Neurophysiological correlates of relatively enhanced local visual search in autistic adolescents. *Neuroimage*, 35(1):283–291.

Manning, C., Kilner, J., Neil, L., Karaminis, T., and Pellicano, E. (2016). Children on the autism spectrum update their behaviour in response to a volatile environment. *Developmental science*.

Mareschal, I., Calder, A. J., and Clifford, C. W. (2013). Humans have an expectation that gaze is directed toward them. *Current Biology*, 23(8):717–721.

Markram, K. and Markram, H. (2010). The intense world theory–a unifying theory of the neurobiology of autism. *Frontiers in human neuroscience*, 4:224.

Marshall, L., Mathys, C., Ruge, D., de Berker, A. O., Dayan, P., Stephan, K. E., and Bestmann, S. (2016). Pharmacological fingerprints of contextual uncertainty. *PLoS biology*, 14(11).

Masi, A., DeMayo, M. M., Glozier, N., and Guastella, A. J. (2017). An overview of autism spectrum disorder, heterogeneity and treatment options. *Neuroscience bulletin*, 33(2):183–193.

Mason, O. J. (2015). The assessment of schizotypy and its clinical relevance. *Schizophrenia bulletin*, 41(suppl 2):S374–S385.

Mathias, S. R., Knowles, E. E., Barrett, J., Leach, O., Buccheri, S., Beetham, T., Blangero, J., Poldrack, R. A., and Glahn, D. C. (2017). The processing-speed impairment in psychosis is more than just accelerated aging. *Schizophrenia bulletin*, 43(4):814–823.

Mayer, J. L. (2017). The relationship between autistic traits and atypical sensory functioning in neurotypical and asd adults: A spectrum approach. *Journal of autism and developmental disorders*, 47(2):316–327.

McCleery, J. P., Allman, E., Carver, L. J., and Dobkins, K. R. (2007). Abnormal magnocellular pathway visual processing in infants at risk for autism. *Biological psychiatry*, 62(9):1007–1014.

McGrath, J., Saha, S., Chant, D., and Welham, J. (2008). Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiologic reviews*, 30(1):67–76.

McMahon, F. J. (2014). Prediction of treatment outcomes in psychiatry—where do we stand? *Dialogues in clinical neuroscience*, 16(4):455.

Merigan, W., Byrne, C., and Maunsell, J. (1991). Does primate motion perception depend on the magnocellular pathway? *Journal of Neuroscience*, 11(11):3422–3429.

Milne, E., Swettenham, J., Hansen, P., Campbell, R., Jeffries, H., and Plaisted, K. (2002). High motion coherence thresholds in children with autism. *Journal of Child Psychology and Psychiatry*, 43(2):255–263.

Mitchell, P. and Ropar, D. (2004). Visuo-spatial abilities in autism: A review. *Infant and Child Development: An International Journal of Research and Practice*, 13(3):185–198.

Montague, P. R., Dolan, R. J., Friston, K. J., and Dayan, P. (2012). Computational psychiatry. *Trends in cognitive sciences*, 16(1):72–80.

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., and Friston, K. J. (2013). Free energy, precision and learning: the role of cholinergic neuromodulation. *Journal of Neuroscience*, 33(19):8227–8236.

Moritz, S. and Woodward, T. S. (2005). Jumping to conclusions in delusional and non-delusional schizophrenic patients. *British Journal of Clinical Psychology*, 44(2):193–207.

Moseley, P., Fernyhough, C., and Ellison, A. (2013). Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. *Neuroscience & Biobehavioral Reviews*, 37(10):2794–2805.

Mottron, L. and Burack, J. A. (2001). Enhanced perceptual functioning in the development of autism.

Mottron, L., Dawson, M., Soulieres, I., Hubert, B., and Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of autism and developmental disorders*, 36(1):27–43.

Moustafa, A. A., Kéri, S., Somlai, Z., Balsdon, T., Frydecka, D., Misiak, B., and White, C. (2015). Drift diffusion model of reward and punishment learning in schizophrenia: Modeling and experimental data. *Behavioural Brain Research*, 291:147–154.

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J., and Friston, K. J. (2014). Bayesian inferences about the self (and others): A review. *Consciousness and cognition*, 25:67–76.

Mulder, M., Van Maanen, L., and Forstmann, B. (2014). Perceptual decision neurosciences–a model-based review. *Neuroscience*, 277:872–884.

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., and Forstmann, B. U.

(2012). Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience*, 32(7):2335–2343.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological cybernetics*, 66(3):241–251.

Nelson, M., Seal, M., Pantelis, C., and Phillips, L. (2013). Evidence of a dimensional relationship between schizotypy and schizophrenia: a systematic review. *Neuroscience & Biobehavioral Reviews*, 37(3):317–327.

Notredame, C.-E., Pins, D., Deneve, S., and Jardri, R. (2014). What visual illusions teach us about schizophrenia. *Frontiers in integrative neuroscience*, 8.

Olney, J. W. and Farber, N. B. (1995). Glutamate receptor dysfunction and schizophrenia. *Archives of general psychiatry*, 52(12):998–1007.

Ospina, M. B., Seida, J. K., Clark, B., Karkhaneh, M., Hartling, L., Tjosvold, L., Vandermeer, B., and Smith, V. (2008). Behavioural and developmental interventions for autism spectrum disorder: a clinical systematic review. *PloS one*, 3(11):e3755.

O'Callaghan, C., Hall, J. M., Tomassini, A., Muller, A. J., Walpola, I. C., Moustafa, A. A., Shine, J. M., and Lewis, S. J. (2017). Visual hallucinations are characterized by impaired sensory evidence accumulation: Insights from hierarchical drift diffusion modeling in parkinson's disease. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 2(8):680–688.

Palmer, C. J., Lawson, R. P., and Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior.

Palminteri, S., Wyart, V., and Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences*.

Pavăl, D. (2017). A dopamine hypothesis of autism spectrum disorder. *Developmental neuroscience*, 39(5):355–360.

Pell, P. J., Mareschal, I., Calder, A. J., von dem Hagen, E. A., Clifford, C. W., Baron-Cohen, S., and Ewbank, M. P. (2016). Intact priors for gaze direction in adults with high-functioning autism spectrum conditions. *Molecular autism*, 7(1):25.

Pellicano, E. and Burr, D. (2012a). Response to brock: noise and autism. *Trends in cognitive sciences*, 16(12):574–575.

Pellicano, E. and Burr, D. (2012b). When the world becomes 'too real': a bayesian explanation of autistic perception. *Trends in cognitive sciences*, 16(10):504–510.

Pernet, C. R., Wilcox, R., and Rousselet, G. A. (2012). Robust correlation analyses: false positive and power validation using a new open source matlab toolbox. *Frontiers in psychology*, 3.

Pilz, K. S. and Papadaki, D. (2019). An advantage for horizontal motion direction discrimination. *Vision research*, 158:164–172.

Pirrone, A., Dickinson, A., Gomez, R., Stafford, T., and Milne, E. (2017). Understanding perceptual judgment in autism spectrum disorder using the drift diffusion model. *Neuropsychology*, 31(2):173.

Pirrone, A., Wen, W., Li, S., Baker, D. H., and Milne, E. (2018). Autistic traits in the neurotypical population do not predict increased response conservativeness in perceptual decision making. *Perception*, 47(10-11):1081–1096.

Plaisted, K. (2001). *Reduced generalization in autism: An alternative to weak central coherence J.A. Burack, T. Charman, N. Yirmiya, P.R. Zelazo (Eds.), The development of autism: Perspectives from theory and research, Lawrence Erlbaum, Mahwah, NJ.* Routledge.

Plaisted, K. C. (2015). Reduced generalization in autism: An alternative to weak central coherence.

Pouget, A., Beck, J. M., Ma, W. J., and Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature neuroscience*, 16(9):1170–1178.

Powell, G., Jones, C. R., Hedge, C., Charman, T., Happé, F., Simonoff, E., and Sumner, P. (2019). Face processing in autism spectrum disorder re-evaluated through diffusion models. *Neuropsychology*.

Powell, G., Meredith, Z., McMillin, R., and Freeman, T. C. (2016). Bayesian models of individual differences: Combining autistic traits and sensory thresholds to predict motion perception. *Psychological science*, 27(12):1562–1572.

Powers, A. R., Kelley, M., and Corlett, P. R. (2016). Hallucinations as top-down effects on perception. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1(5):393–400.

Powers, A. R., Mathys, C., and Corlett, P. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science*, 357(6351):596–600.

Quattrocki, E. and Friston, K. (2014). Autism, oxytocin and interoception. *Neuroscience & Biobehavioral Reviews*, 47:410–430.

Raine, A. (1991). The spq: a scale for the assessment of schizotypal personality based on dsm-iii-r criteria. *Schizophrenia bulletin*, 17(4):555.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature*.

Rao, R. P. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1):79–87.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, 85(2):59.

Ratcliff, R. (2013). Parameter variability and distributional assumptions in the diffusion model. *Psychological review*, 120(1):281.

Ratcliff, R. and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, 20(4):873–922.

Ratcliff, R., Smith, P. L., Brown, S. D., and McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, 20(4):260–281.

Ratcliff, R., Thapar, A., and McKoon, G. (2006). Aging and individual differences in rapid two-choice decisions. *Psychonomic bulletin & review*, 13(4):626–635.

Ratcliff, R., Thapar, A., and McKoon, G. (2010). Individual differences, aging, and iq in two-choice tasks. *Cognitive psychology*, 60(3):127–157.

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

Rauber, H.-J. and Treue, S. (1998). Reference repulsion when judging the direction of

visual motion. *Perception*, 27(4):393–402.

Rauber, H.-J. and Treue, S. (1999). Revisiting motion repulsion: evidence for a general phenomenon? *Vision research*, 39(19):3187–3196.

Regier, D. A., Narrow, W. E., Kuhl, E. A., and Kupfer, D. J. (2009). The conceptual development of dsm-v. *American Journal of Psychiatry*.

Rigoux, L., Stephan, K. E., Friston, K. J., and Daunizeau, J. (2014). Bayesian model selection for group studies—revisited. *Neuroimage*, 84:971–985.

Ring, H. A., Baron-Cohen, S., Wheelwright, S., Williams, S. C., Brammer, M., Andrew, C., and Bullmore, E. T. (1999). Cerebral correlates of preserved cognitive skills in autism: a functional mri study of embedded figures task performance. *Brain*, 122(7):1305–1315.

Robertson, C. E. and Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, 18(11):671.

Rogers, S. J. and Ozonoff, S. (2005). Annotation: What do we know about sensory dysfunction in autism? a critical review of the empirical evidence. *Journal of Child Psychology and Psychiatry*, 46(12):1255–1268.

Ropar, D. and Mitchell, P. (1999). Are individuals with autism and asperger's syndrome susceptible to visual illusions? *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40(8):1283–1293.

Ropar, D. and Mitchell, P. (2001). Susceptibility to illusions and performance on visuospatial tasks in individuals with autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42(4):539–549.

Rosenberg, A., Patterson, J. S., and Angelaki, D. E. (2015). A computational perspective on autism. *Proceedings of the National Academy of Sciences*, 112(30):9158–9165.

Rosenhan, D. L. (1973). On being sane in insane places. *Science*, 179(4070):250–258.

Rust, J. (1988). The rust inventory of schizotypal cognitions (risc). *Schizophrenia Bulletin*, 14(2):317.

Rust, J. (1989). Handbook of the risc. *Kent, England: Psychological Corporation*.

Rutter, M., Le Couteur, A., and Lord, C. (2003). Autism diagnostic interview-revised. *Los Angeles, CA: Western Psychological Services*, 29:30.

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., and Baron-Cohen, S. (2015). Measuring autistic traits in the general population: a systematic review of the autism-spectrum quotient (aq) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular autism*, 6(1):1.

Sandhu, T. R., Rees, G., and Lawson, R. P. (2020). Preserved low-level visual gain control in autistic adults. *Wellcome Open Research*, 4(208):208.

Schatz, J. (1998). Cognitive processing efficiency in schizophrenia: generalized vs domain specific deficits. *Schizophrenia Research*, 30(1):41–49.

Schmack, K., de Castro, A. G.-C., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J.-D., Heinz, A., Petrovic, P., and Sterzer, P. (2013). Delusions and the role of beliefs in perceptual inference. *Journal of Neuroscience*, 33(34):13701–13712.

Schmack, K., Rothkirch, M., Priller, J., and Sterzer, P. (2017). Enhanced predictive signalling in schizophrenia. *Human brain mapping*, 38(4):1767–1779.

Schmack, K., Schnack, A., Priller, J., and Sterzer, P. (2015). Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. *Schizophrenia Research: Cognition*, 2(2):72–77.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., and 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.

Schmitz, F. and Voss, A. (2012). Decomposing task-switching costs with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1):222.

Schneider, K. A. and Komlos, M. (2008). Attention biases decisions but does not alter appearance. *Journal of vision*, 8(15):3–3.

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., and Friston, K. (2015). The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cerebral cortex*, 25(10):3434–3445.

Seida, J. K., Ospina, M. B., Karkhaneh, M., Hartling, L., Smith, V., and Clark, B. (2009). Systematic reviews of psychosocial interventions for autism: an umbrella review. *Developmental Medicine & Child Neurology*, 51(2):95–104.

Seriès, P. and Seitz, A. (2013). Learning what to expect (in visual perception). *Frontiers in human neuroscience*, 7:668.

Seriès, P., Stocker, A. A., and Simoncelli, E. P. (2009). Is the homunculus "aware" of sensory adaptation? *Neural Computation*, 21(12):3271–3304.

Shadlen, M. N. and Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of the national academy of sciences*, 93(2):628–633.

Shams, L., Ma, W. J., and Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, 16(17):1923–1927.

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., and Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision research*, 49(22):2705–2739.

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., Diamond, S. P., and Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences*, 111(42):15220–15225.

Skewes, J. C. and Gebauer, L. (2016). Brief report: Suboptimal auditory localization in autism spectrum disorder: Support for the bayesian account of sensory symptoms. *Journal of autism and developmental disorders*, 46(7):2539–2547.

Skewes, J. C., Jegindø, E.-M., and Gebauer, L. (2015). Perceptual inference and autistic traits. *Autism*, 19(3):301–307.

Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, 123(4):425.

Smith, P. L. and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in neurosciences*, 27(3):161–168.

Sotiropoulos, G., Seitz, A. R., and Seriès, P. (2011). Changing expectations about speed alters perceived motion direction. *Current Biology*, 21(21):R883–R884.

Spaniol, J., Voss, A., Bowen, H. J., and Grady, C. L. (2011). Motivational incentives

modulate age differences in visual perception. *Psychology and aging*, 26(4):932.

Spanò, G., Peterson, M. A., Nadel, L., Rhoads, C., and Edgin, J. O. (2016). Seeing can be remembering: Interactions between memory and perception in typical and atypical development. *Clinical Psychological Science*, 4(2):254–271.

Speechley, W. J., Whitman, J. C., and Woodward, T. S. (2010). The contribution of hypersalience to the "jumping to conclusions" bias associated with delusions in schizophrenia. *Journal of psychiatry & neuroscience: JPN*, 35(1):7.

Srinivasan, M. V., Laughlin, S. B., and Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 216(1205):427–459.

Stafford, M. R., Jackson, H., Mayo-Wilson, E., Morrison, A. P., and Kendall, T. (2013). Early interventions to prevent psychosis: systematic review and meta-analysis. *Bmj*, 346:f185.

Stephan, K. E., Friston, K. J., and Frith, C. D. (2009a). Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia bulletin*, page sbn176.

Stephan, K. E. and Mathys, C. (2014). Computational approaches to psychiatry. *Current opinion in neurobiology*, 25:85–92.

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., and Friston, K. J. (2009b). Bayesian model selection for group studies. *Neuroimage*, 46(4):1004–1017.

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., and Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological psychiatry*, 84(9):634–643.

Sterzer, P., Frith, C., and Petrovic, P. (2008). Believing is seeing: expectations alter visual awareness. *Current Biology*, 18(16):R697–R698.

Stocker, A. A. and Simoncelli, E. P. (2006a). Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience*, 9(4):578–585.

Stocker, A. A. and Simoncelli, E. P. (2006b). Sensory adaptation within a bayesian framework for perception. In *Advances in neural information processing systems*, pages 1289–1296.

Stuke, H., Stuke, H., Weilnhammer, V. A., and Schmack, K. (2017). Psychotic experiences and overhasty inferences are related to maladaptive learning. *PLoS computational biology*, 13(1).

Stuke, H., Weilnhammer, V. A., Sterzer, P., and Schmack, K. (2019). Delusion proneness is linked to a reduced usage of prior beliefs in perceptual decisions. *Schizophrenia bulletin*, 45(1):80–86.

Summerfield, C. and Koechlin, E. (2010). Economic value biases uncertain perceptual choices in the parietal and prefrontal cortices. *Frontiers in human neuroscience*, 4:208.

Sun, J. and Perona, P. (1998). Where is the sun? *Nature neuroscience*, 1(3):183–184.

Sutherland, A. and Crewther, D. P. (2010). Magnocellular visual evoked potential delay with high autism spectrum quotient yields a neural mechanism for altered perception. *Brain*, 133(7):2089–2097.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., and Stewart-Brown, S. (2007). The warwick-edinburgh mental well-being scale (wemwbs): development and uk validation. *Health and Quality of life Outcomes*, 5(1):1.

Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., Goodyer, I. M., and Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences*, 112(43):13401–13406.

Teufel, C., Subramaniam, N., and Fletcher, P. C. (2013). The role of priors in bayesian models of perception. *Frontiers in computational neuroscience*, 7.

Tibber, M. S., Anderson, E. J., Bobin, T., Antonova, E., Seabright, A., Wright, B., Carlin, P., Shergill, S. S., and Dakin, S. C. (2013). Visual surround suppression in schizophrenia. *Frontiers in psychology*, 4:88.

Tomassini, A., Morgan, M. J., and Solomon, J. A. (2010). Orientation uncertainty reduces perceived obliquity. *Vision research*, 50(5):541–547.

Tulver, K., Aru, J., Rutiku, R., and Bachmann, T. (2019). Individual differences in the effects of priors on perception: a multi-paradigm approach. *Cognition*, 187:167–177.

Turi, M., Burr, D. C., Igliozzi, R., Aagten-Murphy, D., Muratori, F., and Pellicano, E. (2015). Children with autism spectrum disorder show reduced adaptation to number. *Proceedings of the National Academy of Sciences*, 112(25):7868–7872.

Turi, M., Karaminis, T., Pellicano, E., and Burr, D. (2016). No rapid audiovisual recalibration in adults on the autism spectrum. *Scientific reports*, 6:21756.

Utzerath, C., Schmits, I. C., Buitelaar, J., and de Lange, F. P. (2018). Adolescents with autism show typical fmri repetition suppression, but atypical surprise response. *Cortex*, 109:25–34.

Utzerath, C., Schmits, I. C., Kok, P., Buitelaar, J., and de Lange, F. P. (2019). No evidence for altered up-and downregulation of brain activity in visual cortex during illusory shape perception in autism. *Cortex*, 117:247–256.

Valton, V. (2014). Impaired reinforcement learning & bayesian inference in psychotic disorders: from maladaptive decision making to pasyhcosis in schizophrenia (phd thesis). pages 103–131.

Valton, V., Karvelis, P., Richards, K. L., Seitz, A. R., Lawrie, S. M., and Seriès, P. (2019). Acquisition of visual priors and induced hallucinations in chronic schizophrenia. *Brain*, 142(8):2523–2537.

Valton, V., Romaniuk, L., Steele, J. D., Lawrie, S., and Seriès, P. (2017). Comprehensive review: Computational modelling of schizophrenia. *Neuroscience & Biobehavioral Reviews*, 83:631–646.

van Boxtel, J. J. and Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in psychology*, 4.

Van de Cruys, S., de Wit, L., Evers, K., Boets, B., and Wagemans, J. (2013). Weak priors versus overfitting of predictions in autism: Reply to pellicano and burr (tics, 2012). *i-Perception*, 4(2):95–97.

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de Wit, L., and Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in

autism. *Psychological Review*, 121(4):649.

Van de Cruys, S., Vanmarcke, S., Van de Put, I., and Wagemans, J. (2018). The use of prior knowledge for perceptual inference is preserved in asd. *Clinical psychological science*, 6(3):382–393.

van der Gaag, M., Smit, F., Bechdolf, A., French, P., Linszen, D. H., Yung, A. R., McGorry, P., and Cuijpers, P. (2013). Preventing a first episode of psychosis: meta-analysis of randomized controlled prevention trials of 12month and longer-term follow-ups. *Schizophrenia research*, 149(1):56–62.

van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P., and Krabbendam, L. (2009). A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness–persistence–impairment model of psychotic disorder. *Psychological medicine*, 39(2):179–195.

Vollema, M. G. and Ormel, J. (2000). The reliability of the structured interview for schizotypy–revised. *Schizophrenia Bulletin*, 26(3):619.

Vollema, M. G. and van den Bosch, R. J. (1995). The multidimensionality of schizotypy. *Schizophrenia Bulletin*, 21(1):19.

von Helmholtz, H. (1867). *Handbuch der physiologischen Optik*, volume 9. Voss.

Vos, T., Barber, R. M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., Charlson, F., Davis, A., Degenhardt, L., Dicker, D., et al. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the global burden of disease study 2013. *The Lancet*, 386(9995):743–800.

Voss, A., Nagler, M., and Lerche, V. (2013). Diffusion models in experimental psychology. *Experimental psychology*.

Voss, A., Rothermund, K., and Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & cognition*, 32(7):1206–1220.

Wacongne, C. (2016). A predictive coding account of mmn reduction in schizophrenia. *Biological psychology*, 116:68–74.

Warner, R. (2013). *Recovery from schizophrenia: Psychiatry and political economy*. Routledge.

Wei, X.-X. and Stocker, A. A. (2015). A bayesian observer model constrained by efficient coding can explain'anti-bayesian'percepts. *Nature neuroscience*, 18(10):1509.

Weiss, Y., Simoncelli, E. P., and Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature neuroscience*, 5(6):598–604.

WHO (1993). *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*, volume 2. World Health Organization.

Winton-Brown, T. T., Fusar-Poli, P., Ungless, M. A., and Howes, O. D. (2014). Dopaminergic basis of salience dysregulation in psychosis. *Trends in neurosciences*, 37(2):85–94.

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232):1880–1882.

Wood, S. J. (2017). Autism and schizophrenia: one, two or many disorders? *The British Journal of Psychiatry*, 210(4):241–242.

Yahata, N., Kasai, K., and Kawato, M. (2017). Computational neuroscience approach to biomarkers and treatments for mental disorders. *Psychiatry and clinical neurosciences*, 71(4):215–237.

Yang, E., Tadin, D., Glasser, D. M., Hong, S. W., Blake, R., and Park, S. (2013). Visual context processing in schizophrenia. *Clinical psychological science*, 1(1):5–15.

Zaidel, A., Goin-Kochel, R. P., and Angelaki, D. E. (2015). Self-motion perception in autism is compromised by visual noise but integrated optimally across multiple senses. *Proceedings of the National Academy of Sciences*, 112(20):6461–6466.

Zamboni, E., Ledgeway, T., McGraw, P. V., and Schluppeck, D. (2016). Do perceptual biases emerge early or late in visual processing? decision-biases in motion perception. *Proceedings of the Royal Society B: Biological Sciences*, 283(1833):20160263.

Zemel, R. S., Dayan, P., and Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural computation*, 10(2):403–430.