Meta-beliefs about the senses: Cognitive and neural mechanisms

Helen Olawole-Scott

Goldsmiths, University of London

Thesis submitted to the Department of Psychology at Goldsmiths, University of London for the degree of Doctor of Philosophy

I, Helen Olawole-Scott, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

Firstly, I extend my deepest gratitude to my supervisor, Daniel Yon, for taking a chance on me as his first PhD student and guiding me with his seemingly endless wisdom, patience, enthusiasm and positivity. I will be forever grateful for your generosity – not only with coffee and 'word of the week' submissions but also with your time, guidance and encouragement. A special note of thanks goes to Rebecca Chamberlain, who stepped in as my Goldsmiths academic support system with kindness and expertise—especially when Daniel abandoned me on his Parisian adventures.

To the amazing people of MERLIN basement – Mal, Denise and Imogen. Thank you for being a constant source of camaraderie and joy throughout the ups and downs of PhD life. Your hugs, voicenotes (aka podcasts) and well-timed withering glances from across the room made this process infinitely more bearable. The friendships forged at Birkbeck will last a lifetime.

It would be foolish of me to not acknowledge the Core4 team - Deepi, Rachel and Tilly - thank you for being the distraction, solace, and cheerleaders I needed at every turn. Your humour, encouragement, and dogged support have been invaluable. I am beyond lucky to have you in my corner, and I hope you never change.

To my parents, thank you for giving me the space and freedom to pursue my academic dreams while cheering me on every step of the way. Your efforts to understand my research means more to me than words can express. To my brother Peter, I owe a special thank you for your patience in helping me re-learn Pythagoras' theorem and troubleshoot coding errors.

My final acknowledgement is to my partner in crime and greatest supporter, Moyin Olawole-Scott: "thank you" feels wholly inadequate for the sacrifices you've made, kindness you've shown, and the laughter you've shared with me throughout this process. Your unwavering belief in me, your understanding, and your stubborn insistence that I should never give up have carried me through even the toughest days. Your love and support have been my rock, and I couldn't have done this without you.

Some of the work reported in this thesis has been published in the following papers:

Chapter 2: **Olawole-Scott, H.**, & Yon, D. (2023). Expectations about precision bias metacognition and awareness. *Journal of Experimental Psychology: General*, *152*(8), 2177.

Chapter 6: Edwards-Lowe, G., La Chiusa, E., **Olawole-Scott, H.**, & Yon, D. (2024, February 15). Information seeking without metacognition. https://doi.org/10.31234/osf.io/cf4a7

Abstract

To effectively interact with the environment, we must estimate the reliability or 'precision' of what we perceive. Assessing our confidence in our senses depends on metacognitive monitoring mechanisms in the brain, which many think are influenced by contextual information and prior beliefs. However, it remains unclear whether and how our brains generate these 'beliefs about precision', and how they influence confidence, awareness and behaviour. This thesis provides an insight into these questions.

The thesis is arranged as follows. After reviewing the relevant literature in **Chapter 1, Chapter 2** describes empirical and computational work testing the influential theory that the mind is 'Bayesian' – the idea that agents form expectations about precision and use these to guide confidence and awareness. In each experiment, participants acquired probabilistic expectations about the likely strength of upcoming signals, while making confidence (*Exps 1-2*) or subjective visibility ratings (*Exp 3*). Computational modelling (*Exp 4*) revealed that the effects of these expectations on awareness could be well-explained by a predictive learning model that infers the precision (strength) of current signals as a weighted combination of incoming evidence and top-down expectation. These results suggest that agents do not only 'read out' the reliability of information arriving at their senses, but also take into account prior knowledge about how reliable or 'precise' different sources of information are likely to be.

Chapter 3 investigated the neural mechanisms underpinning the formation and use of expectations about precision. After adapting and piloting the paradigm used in Chapter 2 (*Exp 5*) we conducted a 3T functional magnetic resonance imaging (fMRI) experiment to investigate the how sensory uncertainty is represented in the brain (*Exp 6*). Using multivariate pattern decoding, we found representations of sensory uncertainty in the insula, which critically were also modulated by expectations about precision. One possibility is that the insula plays a role in encoding 'precision prediction errors' needed to form the expectations about precision seen in Chapter 2.

In **Chapter 4** we investigated whether the 'expected precision' mechanisms identified in previous chapters generalise from the visual domain to our perception of speech – and how these

mechanisms may be connected to unusual hallucination-like experiences. Across two studies (*Exps 7 & 8*) we find that expectations about precision can similarly bias subjective impressions of spoken voices. Moreover, we find some evidence that those prone to hallucinations relied less on these contextual cues when estimating the reliability of the sensory world. This result suggests that the fundamental cognitive and neural mechanisms identified in previous chapters could be disrupted in psychotic illness.

In **Chapter 5** we investigated how expectations about precision influence one aspect of metacognitive control: the control of evidence accumulation (*Exp 9*). And in **Chapter 6** we investigated how representations of sensory uncertainty more generally control another aspect of metacognitive control: seeking information (*Exps 10 & 11*). **Chapter 7** provides a Discussion.

Contents

List of figures and page numbers

Chapter 1: Introduction

Bayesian models of the mind suggest that successful perception, action and cognition depend on estimating uncertainty. For example, tracking the uncertainty of our perceptual systems allows us to engage in sophisticated forms of monitoring and control. Imagine you are driving your car as the sun begins to set. As the sunlight wanes, the information landing at your senses becomes less reliable, leading to less accurate percepts. Importantly, by tracking these changes in sensory reliability, we can act in ways that optimise perception and action. For example, we might turn on the headlights to make things clearer.

Psychologists and neuroscientists have often thought about this kind of uncertainty monitoring through Bayesian models and the idea of *precision*. Bayesian accounts assume that agents track the uncertainty of their own internal states by tracking the noise or variability in different parts of their cognitive machinery. The concept of *precision* features prominently in Bayesian theories of perception, decision-making, and metacognition.

In this Introduction, I outline some examples of how the idea of precision has been applied in different cognitive domains, before pinpointing a key unanswered question in this domain. In particular, current models assume that agents can form 'beliefs about precision' and these beliefs influence how cognition and behaviour unfolds. Testing this idea is the central focus of this thesis – and the Introduction closes with an outline of the thesis, and how different elements help us to answer this question.

1.1. Precision in Bayesian models of perception

Bayesian models of perception propose that individuals estimate the precision of sensory inputs and use these estimates to prioritize information from the most reliable sources during the combination process (Ernst & Banks, 2002; Yon & Frith, 2021). Ernst and Banks (2002) demonstrated this in a study where participants performed a task judging the height of bars using visual and tactile information. By introducing noise to the visual signals, the researchers showed that participants relied on vision when noise was low but shifted to touch when visual noise increased. A maximum-likelihood model supported these findings, showing that participants

adjusted sensory weights based on the reliability of each channel, favouring the more precise input (i.e., more weight was given to visual signals when noise was low, and vice versa when noise was high).

This account has also been used to explain multisensory illusions such as the ventriloquist effect, where the perceived location of a sound is biased towards concurrent visual information, which leads people to believe that a ventriloquist's voice seems to belong to the puppet (Alais & Burr, 2004). Here researchers demonstrated that when visual stimuli are clear and precisely localised, they tend to dominate auditory stimuli, whereas when the visual stimuli were blurred and less reliable, auditory cues dominate. This process aligns with this the model proposed by Ernst and Banks (2002), where the brain weights sensory inputs based on their relative reliability. This idea of combining multiple sensory signals based on precision or reliability of the independent signals forms the basis of many theories of perception and has since been demonstrated to be a cross-species phenomenon (Fetsch et al., 2012; Sheppard et al., 2013).

Precision also features heavily in models of perceptual decision-making but in this case, instead of combining two streams of sensory information, it has been proposed that bottom-up sensory information is combined with top-down expectations about what we are likely to perceive (de Lange et al., 2018; Yon, 2021; Yon & Frith, 2021). For example, in a desert during a sandstorm our incoming visual information is likely to be ambiguous or imprecise (Yon & Frith, 2021). You can generate more reliable inferences of what shapes are in front of you by incorporating prior knowledge (or expectations) into your perceptual processes (e.g., that shape is more likely to be a camel because camels live in the desert, rather than the shape being a polar bear). The weight given to these two sources of information depends on how reliable or 'precise' these two streams of information are likely to be (Yon & Frith, 2021), with expectations being relied on more when sensory information is unreliable or ambiguous. Such mechanisms of precision-weighted decision-making have been extensively supported by behavioural and neuroimaging data (Kok et al., 2013; Lawson et al., 2021; Pinto et al., 2015).

Mechanisms of precision-weighted inference have also inspired theories of unusual experiences and atypical cognition (Corlett et al., 2019; Davies et al., 2018; Yon & Frith, 2021). For example, a

prominent explanation of hallucinations in psychosis suggests that such experiences arise because patients hold inappropriate beliefs about the relative precision of incoming sensory signals and top-down predictions, leading to a disproportionately strong weight on prior expectations when perceiving the world (the 'Strong Priors' theory - Corlett et al., 2019). If the precision of our senses is deemed imprecise or unreliable, then we may tend to rely on our expectations. However, if these expectations do not match reality, yet are dominating our perceptual processes, this may lead to unusual experiences such as hallucination. Hallucinations can be experienced by healthy individuals, not just those with a clinical diagnosis (Corlett et al., 2019). The 'Strong Priors' theory offers a possible mechanism for atypical cognition not just for those with a clinical diagnosis, but across the spectrum of individuals who encounter such experiences (Corlett et al., 2019; Davies et al., 2018; Powers et al., 2016).

1.2. Precision in Bayesian models of metacognition

In a similar vein, Bayesian models of metacognition have suggested that explicit feelings of confidence about what we are perceiving are generated by reading out the uncertainty or precision in sensory circuits – such that we are more confident when our sensory representations are less noisy (Geurts et al., 2022; Mamassian, 2016). This has been demonstrated in behavioural work by Desender and colleagues (2018) where experimental conditions in a perceptual decision-making task were matched in terms of accuracy but differed in participant's subjective evaluation of their accuracy (i.e., their confidence ratings). Results revealed that greater variability or ambiguity in sensory evidence not only impaired performance accuracy but also led to a disproportionately larger decrease in subjective confidence ratings, above and beyond that seen in accuracy (see also Boldt et al., 2017).

Geurts and colleagues (2022) build on these findings with their neuroimaging work. Using psychophysics and functional magnetic resonance imaging (fMRI), they were able to decode probability distributions from neural population activity in the human visual cortex and found that subjective confidence correlated with the precision of the decoded distributions, with higher confidence levels being reported when sensory evidence was more precise. Taken together, these studies suggest that sensory precision tracking is an important mechanism in metacognitive

judgements. However, some researchers argue that the computations underlying confidence are more complex than a simple 'read out' of evidence strength (Aitchison et al., 2015; Fleming & Daw, 2017; Meyniel et al., 2015; Sanders et al., 2016).

1.3. An untested assumption – beliefs about precision?

An important shift in contemporary Bayesian models is the idea that precision is not estimated on incoming evidence alone. Recent accounts also assume that agents form probabilistic beliefs about how precise information is *likely* to be, and these expectations are incorporated into precision estimates (Friston, 2018). Allowing precision to decouple from momentary reality in this way has allowed researchers to develop a myriad of explanations for diverse aspects of experience and awareness. For example, the Strong Priors theory, based in Bayes theorem, suggests that inappropriately strong beliefs about the imprecision of sensory signals (relative to expectations) could lead to unusual and distressing experiences like hallucinations (Corlett et al., 2019).

Forming beliefs about precision would help agents to estimate uncertainty – which may often be difficult to compute (Yon & Frith, 2021). Combining incoming evidence with *expectations* about precision based on past experience could optimise metacognitive monitoring of perception. For example, we may *expect* based on past experience that putting on our glasses will improve the fidelity of incoming visual signals. Since this expectation often comes true, incorporating this prior knowledge into our beliefs will improve our estimates of perceptual precision. However, while combining expectations and incoming signals to estimate precision will usually be adaptive, such a process will also lead to errors when expectation and reality diverge (Figure 1.1). For example, if we mistakenly leave the house with an old pair of glasses that have the wrong prescription, we may believe that putting on our glasses will improve perceptual precision more than it truly does – leading us to overconfidence in our perceptual abilities, with potentially serious consequences as we take our car for a spin.

This idea of *expected precision* has become increasingly embedded in theoretical models of the Bayesian brain. For example, recent models of hierarchical predictive coding suggest that our brain

also entertains a kind of ‘shadow hierarchy’ alongside the primary information streams – with separate neural populations encoding beliefs about the precision of ascending evidence and descending predictions at different hierarchical levels (Friston, 2018). Indeed, computational models based on these frameworks have relied on the concept of expected precision to explain perception (Kanai et al., 2015) and to model false perceptual inferences (Parr et al., 2018). However, while we can potentially explain various aspects of perception and metacognition by assuming agents form beliefs about precision, we do not currently know how or whether expectations about perceptual precision are actually formed. There is extant evidence that agents can predict their decision confidence in a variety of settings (e.g., Boldt et al., 2019; Daniel & Pollmann, 2012; Fleming et al., 2016; Guggenmos et al., 2016), but this does not necessarily entail that agents track or form predictions about the reliability or precision of incoming signals. Without evidence that precision is indeed learned and inferred, it may be premature to use this framework to explain diverse aspects of cognition in health and disease.

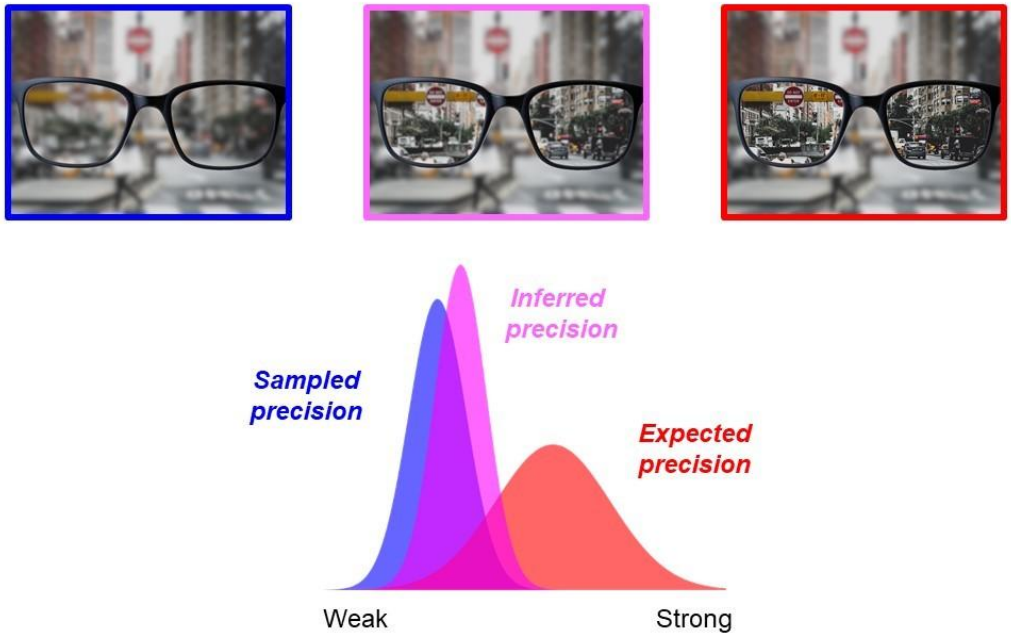


Figure 1.1 – Expectations bias precision estimation: Contemporary Bayesian models suggest that we estimate the precision of our senses by combining incoming evidence with prior expectations about how reliable signals are likely to be. This is usually a good idea but could lead to biases when expectation and reality diverge. For example, if we pick up the wrong pair of glasses, we may expect our vision to improve (red) but the actual signals sampled by vision may remain noisy and imprecise (blue). If we combine this expectation and evidence, we may thus erroneously infer that our vision is more reliable than it truly is (pink).

1.4. Thesis outline

This thesis provides an insight into these questions. First, **Chapter 2** describes empirical and computational investigations testing the influential idea of 'expected precision'. Here we describe novel perceptual decision-making tasks, where cues manipulate an observer's expectations about precision (signal strength). This task or variations of it are employed throughout the experiments of this thesis. In this chapter, I describe how these expectations bias perceptual confidence (Experiments 1 & 2) and subjective visibility ratings (Experiment 3) such that observers were more confident, and stimuli appeared more vivid, when stronger signals were expected. We find this bias in subjective awareness can be captured by a computational model (Study 4) which assumes that

agents form expectations about the signals they are likely to encounter in different contexts and infer the strength (precision) of sensory signals by combining these expectations and incoming evidence from the senses. Taken together, these results support the idea that we estimate the precision of our senses by combining current sensory evidence with expectations about how precise this evidence is likely to be.

Chapter 3 investigates the neural mechanisms underpinning the formation and use of these expectations about precision. After adapting and piloting the paradigm used in Chapter 2 (Experiment 5) we conducted a 3T functional magnetic resonance imaging (fMRI) experiment to investigate the how sensory uncertainty is represented in the brain and where these representations are influenced by expectations (Experiment 6). Using multivariate pattern decoding, we found representations of sensory uncertainty in the insula, which critically were also modulated by expectations about precision. One possible interpretation of our findings is that the insula plays a role in encoding 'precision prediction errors' needed to form the expectations about precision described in Chapter 2.

Chapter 4 investigates whether the 'expected precision' mechanisms identified in previous chapters generalise from the visual domain to our perception of speech – and how these mechanisms may be connected to unusual hallucination-like experiences. Across two studies (Experiments 7 & 8) we find that expectations about precision can similarly bias subjective impressions of spoken voices. In these experiments we also tested the theory put forward by Corlett and colleagues (2019), that inappropriately strong beliefs about the imprecision of sensory signals (relative to expectations) could lead to unusual and distressing experiences like hallucinations. We find some evidence that those prone to hallucinations relied less on contextual cues when estimating the reliability of the sensory world. This result suggests that the fundamental cognitive and neural mechanisms identified in previous chapters could be disrupted in psychotic illness.

While preceding chapters demonstrate that expectations about precision bias participant's metacognitive monitoring (confidence) of perceptual decisions, Chapters 5 and 6 explores whether such expectations can also influence metacognitive control behaviours (i.e., the implementation of

behavioural strategies intending to optimise cognitive performance). In **Chapter 5** (Experiment 9) we investigated how the metacognitive control behaviour of evidence accumulation (i.e., time taken to sample ongoing sensory signals), was influenced by expectations about precision. This experiment found marginal support for the idea that we slow down to accumulate more information when we expect the sensory world to be more ambiguous– however, this difference did not reach statistical significance.

In **Chapter 6** I describe some investigations of how sensory uncertainty itself (rather than expected precision) may influence another form of metacognitive control, information seeking (Experiments 10 and 11). In these experiments participants were offered the chance to have a secondary look at the visual stimulus before committing to their final choice. Results revealed that sensory uncertainty drives information seeking behaviour, in that participants were significantly more likely to opt for a second look at the stimulus when faced with ambiguous sensory evidence compared to strong sensory evidence – even when these differences in ambiguity are separated from decision difficulty.

Chapter 7 provides a discussion of the important themes from each chapter. Overall, our results provide strong support for influential Bayesian models of cognition, showing that agents combine incoming evidence with prior expectations to estimate the precision of their senses (both vision and hearing) and we identify computational and neural mechanisms that elucidate how these processes take place. While there remain some unanswered questions about the role of these expectations in metacognitive control behaviours, overall, this thesis provides support for a widespread but untested tenet of influential Bayesian models of metacognition. It reveals that expectations about precision play an important role in shaping how the sensory world appears and how much we trust our senses.

Chapter 2: Expectations about precision bias metacognition and awareness

2.1. Introduction

Here we investigate whether agents form beliefs about the reliability of incoming sensory signals, and whether these beliefs influence perceptual metacognition in the way that contemporary Bayesian models propose. In particular, we tested the idea that agents form probabilistic beliefs about how precise information is *likely* to be, and that these expectations are incorporated into precision estimates (Friston, 2018).

Participants completed a perceptual decision-making task, judging the direction of moving dots. Crucially, probabilistic cues manipulated expectations about signal strength across trials, such that observers could expect motion clouds to be *strong* or *weak*. To pre-empt our results, across three experiments we found that these expectations biased perceptual confidence (Experiments 1 & 2) and subjective visibility ratings (Experiment 3) such that observers were more confident, and stimuli appeared more vivid, when stronger signals were expected. We find this bias in subjective awareness can be captured by a computational model which assumes that agents form expectations about the signals they are likely to encounter in different contexts and infer the strength (precision) of sensory signals by combining these expectations and incoming evidence from the senses.

Taken together, these results support the idea that we estimate the precision of our senses by combining current sensory evidence with expectations about how precise this evidence is likely to be. This provides support for a widespread but untested tenet of influential Bayesian models of metacognition: revealing that expectations about precision influence how the sensory world appears and how much we trust our senses.

2.2. Experiment 1

Experiment 1 tested whether participants form expectations about the reliability of incoming signals, and how these beliefs influence perceptual metacognition. Participants completed a perceptual decision-making task, judging whether clouds of dots moved left or right. Importantly,

probabilistic cues signalled whether sensory signals would likely be strong or weak. We probed how expectations established by these cues biased perceptual confidence.

2.2.1. Methods

Participants

Thirty-four participants (21 female, 13 male, mean age= 33.9 years, SD= 8.45) were recruited via Prolific. All participants reported normal or corrected vision and no history of psychiatric or neurological illness. This sample size was selected to provide 80% power to detect at least a medium-sized effect (Cohen's dz= 0.5). This value was not explicitly guided by prior work (unlike Experiments 2 and 3). All experiments were approved by the Research Ethics Committee at Goldsmiths, University of London.

Participants who failed to complete at least 90% of trials across the training and test phase were excluded. Participants were considered outliers if their individual effects (i.e., condition-wise differences in accuracy or confidence) were >2.5SDs away from the sample mean. We identified outliers based on participant's condition-wise effects (rather than specific condition means or trial-level data). Outliers were winsorized to values 2.5 SDs away from the mean for inferential statistics, rather than adjusting raw datapoints. The same was true for all experiments. For Experiment 1 no participants were excluded and only one condition-wise effect for confidence scores was winsorized. Data patterns and their statistical significance were unchanged by this adjustment.

Procedure

Participants completed an online perceptual decision-making task programmed in PsychoPy (Peirce et al., 2019), discriminating patterns of moving dots and reporting confidence in their perceptual choices (see Chapter 1, figure 1). Each trial began with a fixation cross (500 ms) before the presentation of a dot motion stimulus (300 ms). In each motion cloud, a proportion of dots was programmed to move coherently left or coherently right, while the remaining dots moved in random

directions. After a blank screen (700 ms) participants gave a combined report of their perceptual decision (left or right) and confidence level (confident or guess) on a four-point scale.

Critically, probabilistic cues (colours) allowed observers to form expectations about the strength of motion signals on each trial, allowing us to investigate how such expectations bias perceptual confidence. For example, for a given observer when fixation cross and stimulus dots appeared in green, motion clouds were likely to have low coherence (i.e., 4% motion coherence - weak signals, see Figure 2.1). In contrast, when stimuli appeared in blue, motion clouds were likely to have high coherence (i.e., 52% motion coherence - strong signals). Colour mappings were counterbalanced across participants and participants were not explicitly informed about the association between the probabilistic cues and signal strength.

The experiment comprised 560 trials (see Figure 2.1). The first 160 trials acted as an initial training phase to establish expectations about colour cues. Here participants experienced perfectly deterministic mappings between colour and coherence e.g., every blue stimulus was programmed to be a strong signal, and every green stimulus was programmed to be a weak signal.

Participants then completed a 400-trial test phase. Half of these trials were identical to what participants experienced during training, where the colour cues were followed by the predicted signal strength (e.g., the colour cue associated with strong signals was followed by objectively strong motion). However, our key trials of interest in this phase were the remaining half of *medium probe* trials. On these trials, participants received the same colour cues but received an objective perceptual signal of medium strength (28% motion coherence). Given objective signal strength is identical on these trials, any differences in objective accuracy or subjective confidence on these trials must reflect effects of expectations about signal strength.

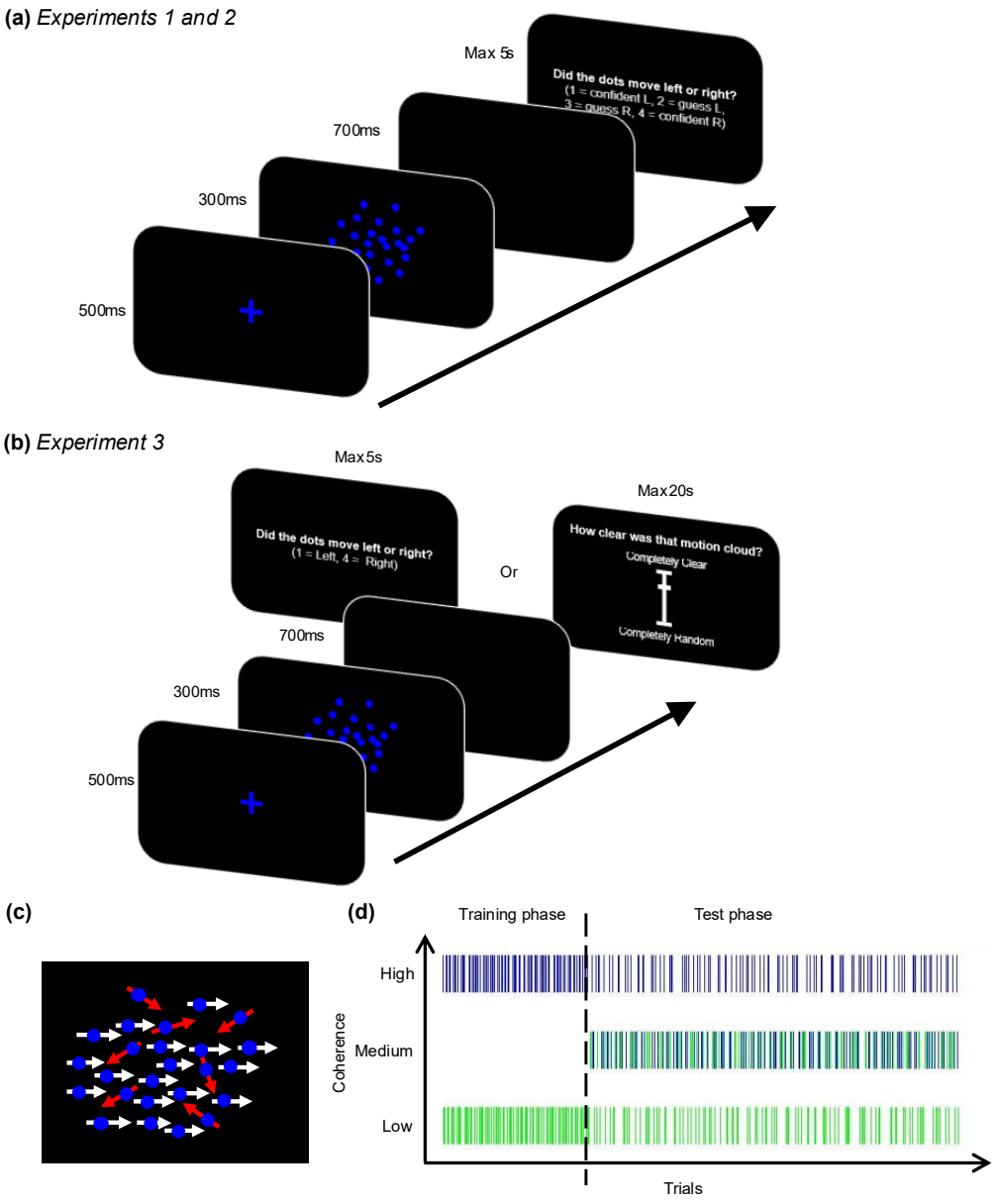


Figure 2.1 – Experimental task: (a) Participants completed a motion perception task, judging the direction of brief motion clouds and reporting confidence in their decision. Colour cues manipulated expectations about the strength of motion patterns for each trial, e.g., if stimuli were blue participants could expect high motion coherence. **(b)** On *medium probe trials* in Experiment 3 the perceptual decision was replaced by a visibility scale. **(c)** Illustration of 'motion coherence': in each stimulus, a proportion of dots was programmed to move left or right (white arrows) while the remainder of dots moved in random directions (red arrows). Manipulating the proportion of coherent dots changes the strength of the motion signal. **(d)** Example timecourse of trials across the experiment: The training phase consisted of perfectly deterministic mappings between colour and coherence. During the test phase half of the trials were identical to those shown during training, whereas the other half paired the same colour cues with objective perceptual signals of medium strength.

2.2.2. Results

We investigated how actual and expected precision altered perception and metacognition by computing measures of objective and subjective performance from perceptual choices and confidence ratings. Objective perceptual sensitivity was measured by calculating the proportion of correct decisions (accuracy) and d'. We also looked at reaction times on these trials. To capture subjective aspects of metacognition, we calculated 'confidence level' i.e., the proportion of decisions participants report with high rather than low confidence. We also calculated meta-d' and Mratio as complementary measures of metacognitive sensitivity and efficiency – measuring how closely subjective confidence ratings ('confident' or 'guess') track decision accuracy, and whether this changes while controlling for differences in task performance (Maniscalco & Lau, 2012). We computed d', meta-d' and Mratio using the non-hierarchical variant of the HMeta-d toolbox (Fleming, 2017). Inferential tests used an alpha level of .05, and non-significant results were qualified with equivalent Bayesian analyses. These yielded Bayes Factors (BF) that quantified evidence for an effect (H_1) over evidence for the null (H_0). Conventionally BF_{10} <.33 denotes moderate evidence in support of a null effect.

First, we compared perception and metacognition on trials where motion signals were objectively stronger or weaker. Unsurprisingly, perceptual decisions were more accurate when signals were objectively stronger (mean accuracy = 0.939, mean d' = 3.672) than when they were objectively weaker (mean accuracy = 0.564, mean d' = 0.349; accuracy – *t(33)= 21.606, p<.001, dz=3.705; d'- t(33)= 18.061, p<.001, dz= 3.097*). Participants also reported higher confidence in perceptual decisions when objective signal strength was strong (mean = 0.843) compared to when signals were objectively weak (mean = 0.341, *t(33) = 11.200, p<.001, dz = 1.921*).

However, our key question concerns how *expected* precision alters perception and metacognition. This can be evaluated by comparing our test trials where participants *expect* strong or weak signals – but receive objectively identical medium coherence stimuli. These comparisons revealed that participants were more confident in their decisions on *expect strong* trials – mean (SEM) = 0.71 (0.042) – than *expect weak* trials (mean (SEM) = 0.64 (0.045), *t(33)= 3.015, p= .005, dz= 0.517*, see Figure 2.2).

Importantly, this difference in confidence arose even though objective perceptual accuracy and d' scores did not significantly differ between these conditions (accuracy: *t(33)= 0.913, p= .368, dz= 0.156*, BF_{10} = 0.270; d': *t(33)= 0.194, p= .847, dz= 0.033*, BF_{10} = 0.187, respectively, see Figure 2.3). There was also no difference in reaction time (*t(33)= 1.257, p= .217, dz=0.216,* BF_{10} = 0.378). There was also no significant difference in meta-d', nor Mratio, between conditions (meta-d': mean for *expect weak* trials = 2.05, mean for *expect strong* trials = 2.312, *t(33)= 1.544, p=.132, dz=0.265*, BF_{10} = 0.539; Mratio: mean for *expect weak* trials = 0.94, mean for *expect strong* trials = 0.962, *t(33)= 0.146, p= .884, dz=0.025*, BF_{10} = 0.186) – suggesting that expectations about precision induce a metacognitive bias, rather than altering the discriminability of introspective states.

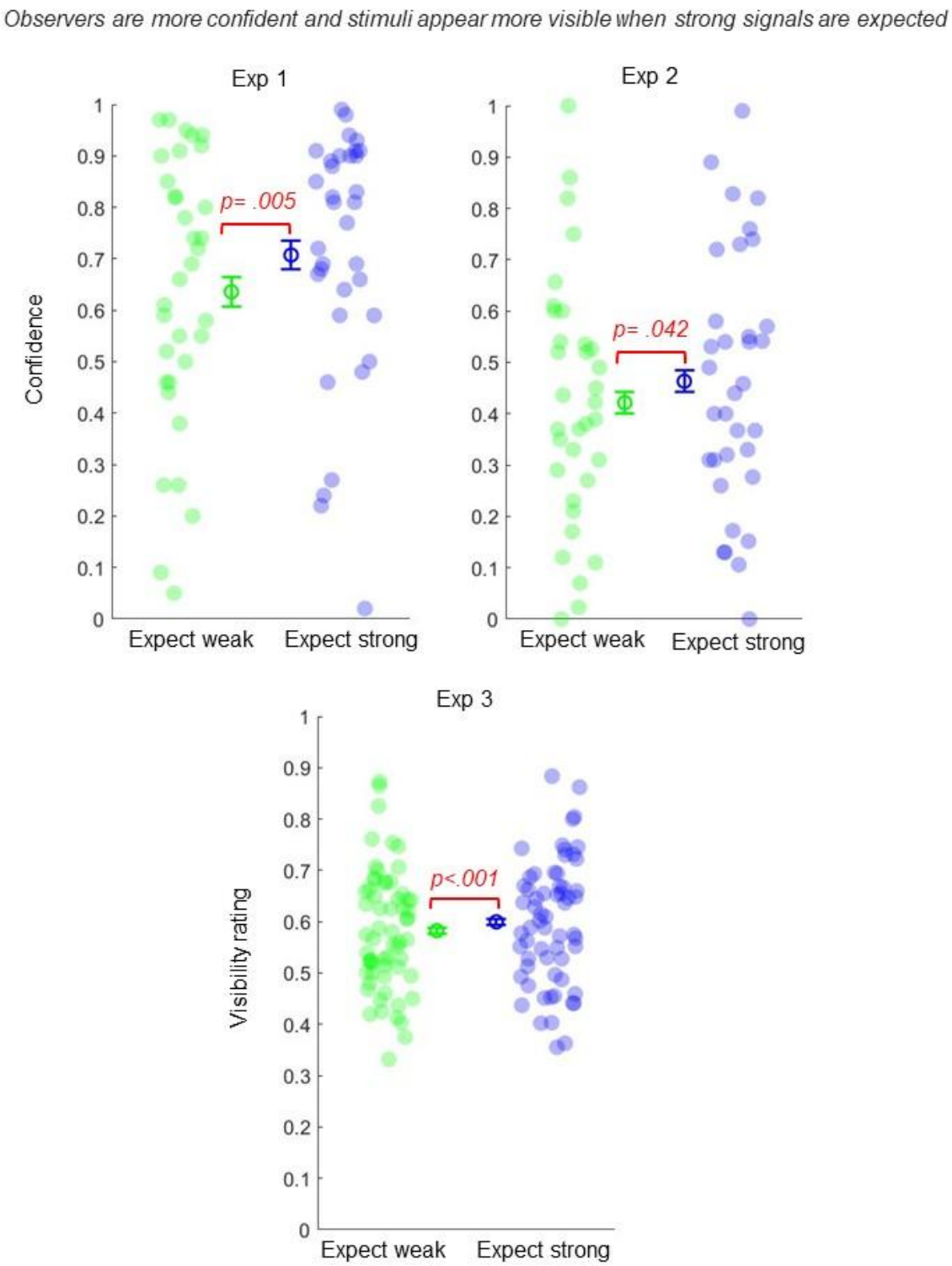


Figure 2.2 - **Expected precision alters metacognition and awareness***:* Participants reported significantly higher confidence (Experiments 1 and 2) and higher visibility ratings (Experiment 3) on 'expect strong' trials. Error bars represent 95% within-subject confidence intervals.

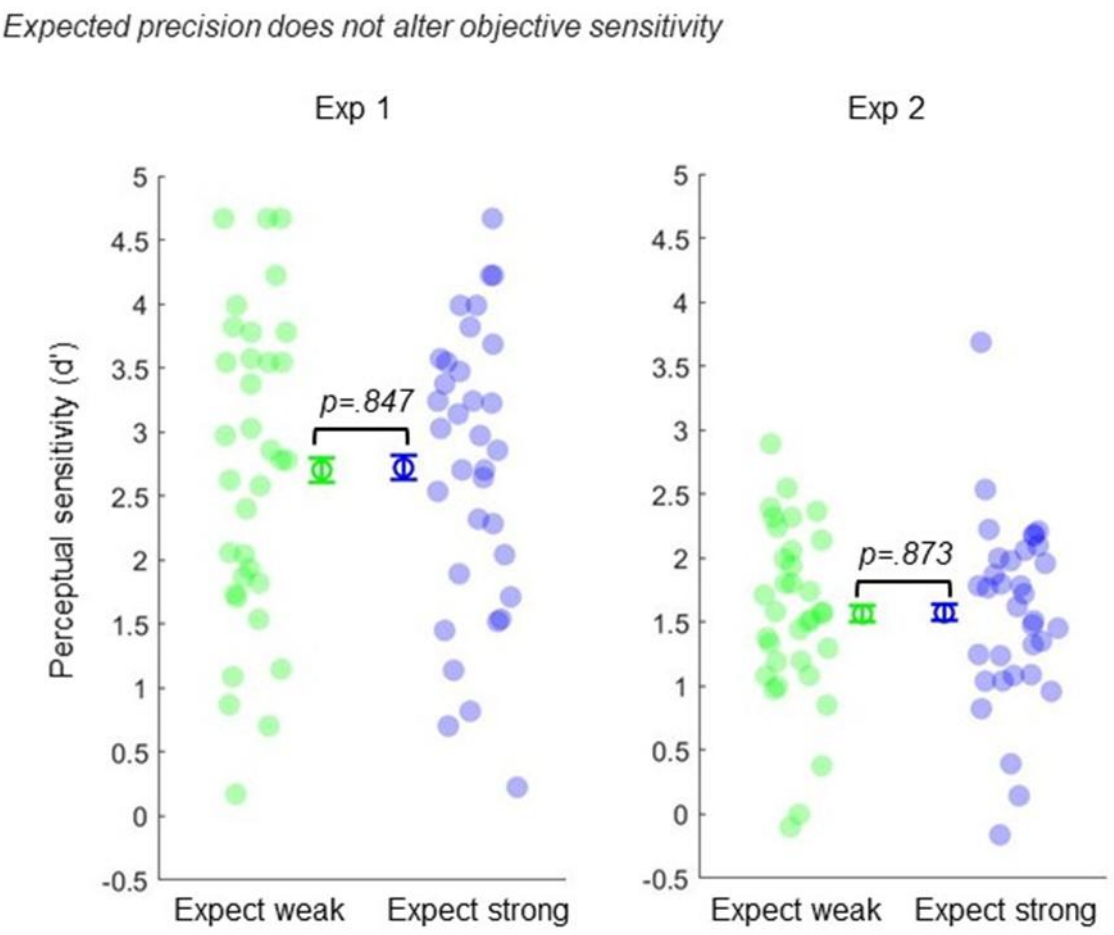


Figure 2.3 – **Expected precision does not alter objective sensitivity***:* Perceptual sensitivity (d') was unaffected by probabilistic cues. Error bars represent 95% within-subject confidence intervals.

2.2.3. Discussion

Experiment 1 suggests that expectations about signal strength bias perceptual metacognition, such that perceptual confidence is exaggerated when participants expect more precise (i.e., high coherence) motion signals – even if such strong signals do not actually ensue. This is consistent with Bayesian theories that suggest we form expectations about the precision of sensory signals (Friston, 2018), which in turn shape beliefs about the reliability of the senses (Yon & Frith, 2021).

2.3. Experiment 2

Experiment 1 found evidence consistent with the Bayesian idea that expectations about signal strength bias perceptual metacognition. This inference was based on the fact that perceptual confidence differed for objectively identical stimuli according to whether the observer *expected* a strong or weak signal, while perceptual and metacognitive sensitivity remained unchanged.

In Experiment 1 these medium 'test' stimuli were chosen as the midpoint of coherence (28%) between weak (4%) and strong (52%) signals participants experienced throughout the task. While this makes the medium test stimuli the objective intermediate point between the signals, in Experiment 1 such stimuli were found to not be intermediate in terms of decision difficulty. In particular, accuracy on medium test trials (mean = 0.872) was more similar to decision accuracy with strong (mean = 0.939) rather than weak signals (mean = 0.564).

This discrepancy is potentially important for understanding the underlying mechanism at play in Experiment 1. For example, it is possible that expecting strong signals actually improves the *sensitivity* of metacognition, such that participants are better able to detect their correct decisions, rather than directly inducing a confidence bias (as has been suggested in prior work - Sherman & Seth, 2021). When accuracy is near ceiling, an improvement in metacognitive sensitivity may appear to induce a bias in overall confidence – since accurate insight will lead to higher confidence ratings.

This alternative explanation seems unlikely given that Experiment 1 found expectations biased perceptual confidence but did not significantly alter metacognitive sensitivity (meta-d). However, to ensure the reliability of these effects and to rule out this alternative explanation we ran Experiment 2. Experiment 2 was a pre-registered replication of Experiment 1 with one key change: the coherence of *medium probe* trials was lowered to ensure participants would no longer approach ceiling on decision accuracy. If expectations about precision directly bias confidence, Experiment 2 should replicate the findings of Experiment 1 – finding expectations about signal precision induce a bias in confidence ratings but leave metacognitive sensitivity unaffected.

2.3.1. Methods

Participants

For Experiment 2, thirty-four new participants (15 female, 19 male, mean age = 37.3, SD = 9.28) were recruited via Prolific. This sample size was selected to provide 80% power to detect effects at least as large as those found in Experiment 1 (effect of expectation on confidence level - dz = 0.517). The same exclusion and outlier identification procedures were used as in Experiment 1. No participants were excluded, and winsorization was applied to one participant's condition-wise effect in the confidence level comparison – though this adjustment did not affect data patterns or their significance.

Procedure and Paradigm

Experiment 2 used the same procedure and paradigm as Experiment 1, except the coherence level of the middle signal strength trials was decreased from 28% to 16% motion coherence.

2.3.2. Results

The same measures from Experiment 1 (accuracy, d', reaction times, confidence, meta-d' and Mratio) and statistical analyses were also employed for Experiment 2.

Experiment 2 replicated the results of Experiment 1. Again, participants were more accurate when signals were objectively stronger (mean accuracy = 0.945, mean d' = 3.761) than when they were objectively weaker (mean accuracy = 0.567, mean d' = 0.339; accuracy – $t(33)= 20.444, p<.001, dz= 3.506$; d'- $t(33)=18.177, p<.001, dz= 3.117$). Participants also reported higher confidence in perceptual decisions when objective signal strength was strong (mean = 0.800) compared to when signals were weak (mean = 0.294, $t(33)= 10.332, p<.001, dz = 1.772$).

More importantly, participants reported higher confidence on *expect strong* trials (mean (SEM) = 0.46 (0.042) compared to *expect weak* trials (mean (SEM) = 0.42 (0.041), $t(33)= 2.114, p=.042, dz= 0.362$, see Figure 2.2).

Again, these differences in confidence were obtained even though there was no significant difference in accuracy (*t(33)= 0.118, p=.907, dz= 0.020*, BF_{10} = 0.185) or d' between conditions (*t(33)=0.161, p= .873, dz= 0.028*, BF_{10} = 0.186; see Figure 2.3). There were also no differences in reaction times (*t(33)= 1.146, p= .260, dz= 0.196*, BF_{10} = 0.335). Critically, while this change in confidence level was replicated, expectations about signal precision had no significant effect on metacognitive sensitivity (meta-d: mean for *expect weak* trials = 1.121, mean for *expect strong* trials = 1.092, *t(33)= 0.261, p= .796, dz = 0.045*, BF_{10} = 0.190), nor metacognitive efficiency (Mratio: mean for *expect weak* trials = -69.349, mean for *expect strong* trials = 24.659, *t(33)= 0.998, p= .326, dz= 0.171*, BF_{10} = 0.291).

2.3.3. Discussion

Experiment 2 replicated the results of Experiment 1 – again finding that observers are biased to feel more confident in perceptual choices when stronger signals are expected. These effects are consistent with Bayesian models that assume agents form beliefs about the precision of incoming signals and use these expectations to guide perceptual metacognition. Importantly, Experiment 2 also rules out the possibility that these effects arise because of changes in metacognitive sensitivity rather than bias.

However, this is not the only interpretation. An alternative possibility is that this bias in confidence arises because agents form expectations about their *performance* in different contexts, rather than expectations about the precision of evidence per se. For example, previous work has found that agents readily form beliefs about the difficulty of different tasks even in the absence of explicit feedback – and can use these global performance estimates to guide decisions about which goals to pursue (Rouault et al., 2019). Indeed, recent results suggest that these kinds of expectations about confidence and task difficulty may also directly bias prospective and retrospective decision confidence (Boldt et al., 2019; Van Marcke et al., 2022). Under this alternative way of thinking, effects in our experiments may be generated by metacognitive mechanisms which track the fact that decisions tend to be more accurate in one colour context than another. Learning about the probability of being correct could also bias decision confidence (Fleming & Daw, 2017), even if

agents are not learning or forming expectations about the clarity or precision of incoming signals – but simply learn that they feel more confident in some contexts rather than others, without learning why.

To evaluate this alternative possibility, we ran Experiment 3 – testing more directly whether observers acquire expectations about precision, and whether these expectations shape inferences observers make about the strength and clarity of incoming sensory signals.

2.4. Experiment 3

Experiments 1 and 2 found that participants reported higher confidence in perceptual decisions when expecting strong signals. Such biases could be driven by changes in apparent signal strength – such that when the participant is expecting a stronger signal, they overestimate the precision of incoming sensory information, leading to exaggerated feelings of confidence. However, as noted above, this effect could also reflect participants forming a confidence bias which is unrelated to signal precision – for example, learning that decisions tend to be easier in the blue context rather than the green context, without tracking signal strength to learn this.

Experiment 3 was conducted to determine whether learning in our paradigm biases the apparent reliability of perceptual signals, rather than inducing a generic confidence bias. To this end, we replaced confidence ratings with a more direct assay of apparent signal strength – the subjective clarity of the visual motion.

2.4.1. Methods

Participants

For Experiment 3, sixty-two new participants (36 female, 26 male, mean age = 35.4, SD = 7.52) were recruited via Prolific, using the same selection criteria as Experiment 1 and 2. The sample size was chosen to provide 80% power to detect effects at least as large as those seen on confidence level in Experiment 2 (Cohen's dz = 0.362). The experiment used the same exclusion

and outlier identification criteria as Experiments 1 and 2. No participants were excluded. Winsorization was applied to one participant's condition-wise effect (in the visibility rating comparison), but this did not alter statistical patterns or their significance.

Procedure and Paradigm

Participants completed the same task used in Experiment 1 with two critical changes. The metacognitive report was removed entirely, such that participants only reported motion direction (left or right) and never rated decision confidence. On the critical *medium probe* trials in the test phase, participants did not make judgements about motion direction. Instead, a visibility scale appeared post-stimulus, asking participants to judge "how clear was that motion cloud?" on a continuous scale ranging from 'completely random' to 'completely clear' (see Figure 2). Ratings from this scale were used as an index of subjective awareness, providing an assay of how clear (or 'precise') visual signals appeared. Importantly, removing the perceptual decision on these trials means that participants must make an estimate about the sampled signal, rather than estimating the correctness of an explicit choice.

The overall structure of the experiment remained the same as Experiments 1 and 2, except the number of training phase trials was increased to 200. During the test phase, the visibility scale was displayed on all *medium probe* trials instead of the perceptual decision. To prevent participants from learning that the visibility scale was only presented on trials where the objective signal strength was truly intermediate, the scale was also presented on 10% of the high and low coherence trials in the test phase.

2.4.2. Results

As in Experiments 1 and 2, perceptual decisions were more accurate when signals were objectively stronger (mean accuracy = 0.972, mean d' = 3.911) compared to when they were objectively weaker (mean accuracy = 0.562, mean d' = 0.326; accuracy – $t(61)= 55.718, p<.001, dz=7.076$; d'- $t(61)= 41.220, p<.001, dz= 5.235$). Participants also reported higher subjective

visibility ratings when objective signal strength was strong (mean = 0.74) compared to when signals were weak (mean = 0.417, *t(61) = 15.236, p<.001, dz= 1.935*).

Critically, analyses also revealed that expectations about precision altered visibility ratings, even when objective signal strength was matched. Observers reported that medium strength stimuli appeared more vivid when stronger signals were expected - mean (SEM) = 0.6 (0.015) – compared to when signals were expected to be weak - mean (SEM) = 0.582 (0.015), *t(62) = 3.673, p<.001, dz= 0.467* (see Figure 2.2).

2.4.3. Discussion

It was unclear from Experiments 1 and 2 whether this learning manipulation causes participants to form expectations about signal strength or expectations about performance confidence. In Experiment 3 participants rated the subjective clarity of motion, rather than reporting decision confidence. Here we found that observers were biased to rate identical motion clouds as seeming clearer when more precise signals were expected. This is consistent with the possibility that our learning manipulation causes observers to form expectations about perceptual precision which alter how strong signals appear to be. These changes in apparent signal quality can also plausibly explain why decision confidence is higher when stronger signals are expected.

2.5. Study 4 - Computational Modelling

Experiment 3 found that stimuli appeared more vivid when the observers expected stronger signals. Such an effect could arise if subjective vividness reflects an inference of signal strength, which observers form by combining the bottom-up sensory evidence with top-down predictions about how strong or 'precise' sensory evidence is likely to be in a given context (Friston, 2018; Yon & Frith, 2021). We used computational modelling to evaluate this possibility.

In our model agents learn and make inferences about the signals they encounter in different contexts throughout the task. On each trial, observers receive a stimulus with a certain signal strength – ranging from completely random motion to completely coherent motion in one direction.

As a first step, we assume the model has access to these stimulus energies trial-by-trial. We estimated the motion energy in each stimulus by calculating the horizontal motion component of each moving dot (given by the cosine of the motion angle). Averaging these motion components across all dots in the display yields an estimate of signal strength bounded between 1 (all dots move coherently in one direction) and 0 (no motion signal at all). While the motion signal present on any given trial is strongly determined by its programmed coherence (i.e., higher coherence clouds tend to have stronger signals), there can still be substantial variability between motion clouds with the same programmed coherence – depending on how the random dots in the cloud behave.

Our model assumes that agents use these samples of motion energy to learn expectations about the likely signal strengths in the two cue contexts (i.e., blue context and green context), which in turn shape estimates about signal strength on a given trial. The model implements this by assuming that an inference of signal strength *(inference*) on trial *t* is produced by computing a weighted average of the sampled sensory evidence (*evidence*) and prior expectation (*prior*), where w_{prior} and $w_{evidence}$ are the respective weights applied to expectations and evidence in this combination:

$$inference_t = w_{prior}(prior_t) + w_{evidence}(evidence_t)$$

$$w_{evidence} = 1 - w_{prior}$$

In this equation there is only one free parameter – w_{prior} – which controls the relative impact that prior expectation and current evidence have on internal estimates of signal strength. If $w_{prior} = 1$, the observer's current belief about signal strength is entirely determined by their previous experience in this context, ignoring the present stimulus entirely. In contrast, if $w_{prior} = 0$ internal beliefs about signal strength are entirely driven by the quality of the current stimulus and past experience is discarded.

Importantly, the model iteratively combines learning and inference, such that once a belief about signal strength has been formed on trial *t* this becomes the new prior for that context on trial *t*+1. This is analogous to iterative Bayesian updating schemes where prior and evidence are combined

at one timepoint to form a posterior, which becomes the new prior for the next timepoint, and so on. Importantly, this means that w_{prior} also effectively acts as a learning rate parameter. For values of w_{prior} closer to 0, expectations for trial $t + 1$ are driven mostly by signals experienced on trial t. In contrast, for values of w_{prior} closer to 1, predictions are more strongly driven by the accumulation of past experiences rather than current evidence. This dual role for the parameter w_{prior} in inference and learning is reminiscent of hierarchical Bayesian models of message passing in the brain, which assume a common parameter simultaneously determines how strongly prior knowledge is weighted when making inferences and how stubborn these prior hypotheses are in the face of new data (Friston, 2018; Yon et al., 2019). Indeed, under certain assumptions this learning model can be shown to be equivalent to models of Bayesian inference, where the combination of prior and evidence is controlled by the (estimated) precision of each information source (see Supplementary Modelling).

This process yields a trajectory of beliefs about signal strength that integrates past experience and current evidence – controlled by the parameter w_{prior} (see Figure 2.4).

The final step in the model turns trial-wise beliefs about signal strength into ratings on the visibility scale. This is achieved by taking the internal inference of signal strength on a given trial and passing this through a logistic function of the form:

$$rating = \frac{1}{1+e^{-(b_{const}+b_{slope}(inference_t))}}$$

This transfer function in the model reflects our assumption that agents form beliefs about signal strength that they communicate in potentially noisy or biased ways. This accords with ideas from metacognition research (Bang et al., 2020; Guggenmos, 2022) or reinforcement learning (Lockwood & Klein-Flügge, 2021) where decisions and actions reflect a noisy transfer of an internal belief into an overt choice. This function produces a continuous rating of motion vividness bounded between 0 (completely random) and 1 (completely clear), controlled by two parameters – b_{slope} and b_{const}. b_{slope} determines the gradient of the function mapping internal estimates of signal strength to visibility ratings – such that higher values indicate a tight mapping between beliefs and ratings, and

lower values indicate a noisier translation from inferences to ratings (*b – beta)*. The b_{const} parameter is a constant value, that captures idiosyncratic biases to give high or low visibility ratings irrespective of current inferences about signal strength.

To investigate whether this three-parameter model could capture empirical patterns seen in Experiment 3, for each participant we simulated belief trajectories for values of w_{prior} ranging between 0 and .999 and subsequently found values of b_{slope} and b_{const} that best predicted the empirical visibility ratings participants gave on medium strength test trials. Maximum likelihood estimation of the logistic transfer function allowed us to identify the combination of best-fitting parameters that minimised the deviance between model and data (i.e., maximised model evidence).

Identifying the best fitting parameters for each participant allows us to simulate how the model behaves in the experiment, and to investigate whether the model reproduces the observed empirical effects. Analysing simulated data in the same way as real data found that the model reproduced the key result of Experiment 3 – yielding higher subjective visibility ratings on medium test trials when stronger rather than weaker signals were expected (t_{61} = 7.366, p<.001, dz = 0.935; see Figure 2.4). Moreover, we found a strong correlation between the size of the empirical effect for each participant and the size of this effect predicted by the model – r = 0.731, p<.001.

Analysing parameter values allowed us to examine which aspects of the model contribute to its ability to reproduce these empirical effects. We found a strong relationship between values of parameter w_{prior} and the empirical effect observed for each participant in Experiment 3 – r = 0.363, p=.004 – suggesting that those participants who showed the largest effects of expectations about signal strength were those the model estimated to be placing the greatest weight on prior knowledge (see Figure 2.4).

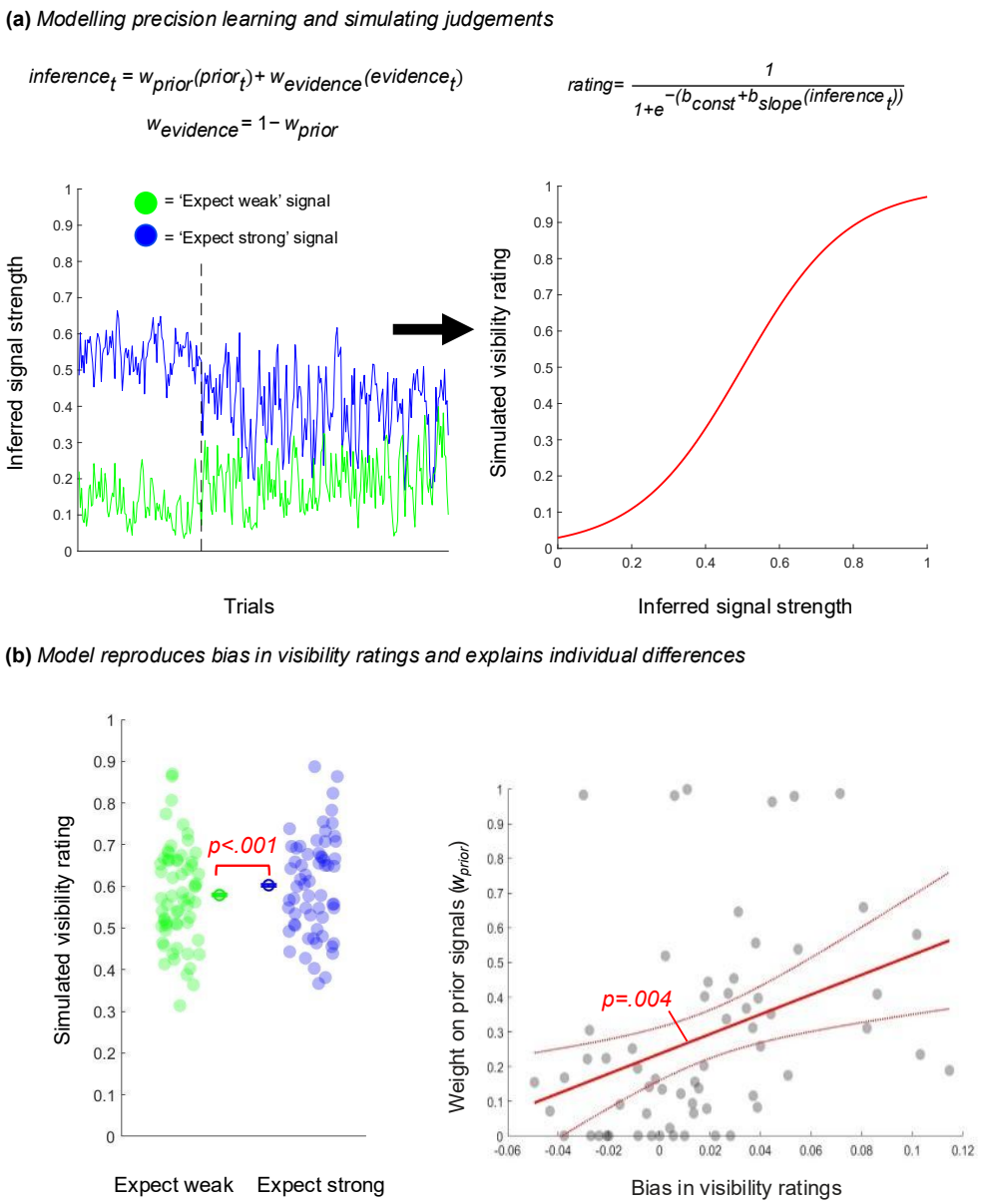


Figure 2.4- Modelling precision learning: (a) Our model assumes that observers form inferences about signal strength by computing a weighted combination of incoming evidence and past experiences. This generates a trajectory of beliefs about signal strength across trials (left). Our model assumes when observers are probed to rate the visibility of a stimulus, they pass this momentary belief through a mapping function to generate a rating (right). Dashed line in left panel denotes transition from training phase to test phase. **(b)** Analysing simulated data in the same way as real data found that the model reproduced the key result of Experiment 3 – the model rates stimuli as being more visible when stronger signals should be expected (left). There was a strong correlation between the model weight on prior experience (w_{prior}) and the empirical bias observed for each participant (higher values = stronger bias). Red lines display best linear fit and confidence bounds.

2.6. General discussion

Influential theories suggest that the mind is *Bayesian* – computing the uncertainty or precision of internal representations to guide perception, action, and cognition. In particular, Bayesian accounts of metacognition propose that we build representations of perceptual confidence by estimating the precision of representations in our sensory systems (Geurts et al., 2022). However, an important recent shift in Bayesian frameworks has been the emerging idea that precision estimates are not simply 'read out' from sensory systems but formed by combining incoming evidence with learned expectations about how reliable sensory evidence is likely to be. This idea has been and continues to be very influential across the cognitive sciences but has not been directly tested (Yon & Frith, 2021). Here we tested this possibility by manipulating participant's expectations about precision (signal strength) and measuring how these altered perceptual metacognition and subjective awareness.

Our results support the idea that agents combine incoming evidence with learned expectations to estimate the precision of sensory information. We found that participants reported higher confidence (Experiments 1 and 2) and more vivid percepts (Experiment 3) when they expected signals to be stronger – even though objective signal strength was identical, and objective perceptual performance remained unchanged. These results were complemented by computational modelling, which revealed such biases could be well-explained by assuming agents infer the precision of sensory signals by combining immediate evidence from their perceptual systems with expectations about how strong signals are likely to be (Friston, 2018; Yon & Frith, 2021).

These results have important implications for our understanding of metacognition and perceptual monitoring. One influential conceptualisation defines metacognitive states as those that represent uncertainty in our overt and covert decisions (Pouget et al., 2016). This distinguishes metacognition from other kinds of meta-representation in the mind and brain (Shea, 2012). In this way of thinking, perceptual precision estimates are metarepresentational, because they represent uncertainty about the perceptual world. But under this definition, they are not strictly 'metacognitive', as they do not directly represent uncertainty in our decisions.

However, even if perceptual precision estimates are not metacognitive in this sense, they can still support important perceptual monitoring functions – allowing observers to estimate the clarity of their senses. Critically, these estimates of perceptual evidence strength can then form an important component of strictly metacognitive computations like decision confidence (Mamassian, 2016). Though the computations underlying confidence are more complex than a simple 'read out' of evidence strength (Aitchison et al., 2015; Fleming & Daw, 2017; Meyniel et al., 2015; Sanders et al., 2016), many models also assume that biases to over- or under-estimate the strength of sensory evidence should also translate into biases in decision confidence - as we have found in the present work.

For example, normative models propose that we estimate the confidence in our perceptual decisions using estimates of the uncertainty in our perceptual circuits (Geurts et al., 2022). However, it is likely to be difficult for systems like the brain to monitor uncertainty based on incoming signals alone (Yon & Frith, 2021). Our results suggest metacognitive mechanisms may finesse this problem by incorporating prior knowledge into these computations – inferring how reliable our senses are by combining current evidence from our perceptual systems with expectations about how precise they are likely to be (Friston, 2018). This will often be adaptive because expectations about precision will often come true. For example, I may *expect* my vision to improve when I put on my glasses, and if I have the right prescription; this expectation is valid. However, relying on expectations may lead to false metacognitive inferences when prediction and reality do not coincide. For example, if I have picked up the wrong pair of glasses, I may expect to see more clearly but actually be more myopic than before. In these cases, relying too heavily on expected precision will lead to overconfidence and maladaptive action based on unreliable evidence.

Our findings demonstrate that agents form expectations about sensory precision which directly alter inferences about signal quality. Previous models (Fleming & Daw, 2017) and experiments (Rouault et al., 2019; Sherman & Seth, 2021) have assumed that agents form metacognitive expectations about task performance – often conceptualised as expecting a high or low probability of being correct. Such ideas gel with computational accounts of metacognition, which define

metacognitive processes as exclusively being those involved in computing the probability that a decision is correct (Pouget et al., 2016). Here, we find expected precision biases perceptual confidence (Exps 1 and 2) but also see these expectations directly alter judgements of signal strength even when no decision is required (Exp 3). These results may suggest an intermediate stage of perceptual monitoring between lower-level perception and higher-level metacognition, where agents compute the strength of incoming signals rather than the accuracy of their decisions per se. Indeed, elegant neuroimaging work has found neural representations encoding the quality or vividness of sensory signals that are distinct from those encoding decision confidence (Bang & Fleming, 2018; Mazor et al., 2022). The results we report here are thus compatible with a view where expected precision alters these mid-level representations of signal quality (Yon & Frith, 2021): directly altering how reliable signals appear to observers, which in turn biases later computations of decision confidence that depend on this information.

Experiment 3 investigated whether agents genuinely form expectations about signal strength, rather than performance confidence alone. This was achieved by asking participants to rate subjective clarity of motion clouds rather than reporting decision confidence. Results showed that observers were biased to rate identical motion clouds as seeming clearer when more precise signals were expected – even when this was probed independently of any 'decision'. Conceptually, it makes sense to distinguish visibility ratings from confidence ratings, since one judgement asks about properties of the stimulus and the other asks about properties of the decision maker (and indeed, the two kinds of ratings often empirically decouple - Davidson et al., 2022; Rausch & Zehetleitner, 2016; Skewes et al., 2021). However, it remains possible in principle that covert metacognitive processes may influence visibility ratings - such that a participant judges a stimulus is more visible because they judge they could (hypothetically) make an accurate decision about it if probed. Future work could assess this possibility by creating paradigms where confidence and visibility are more strongly decorrelated. This could be achieved by altering the base rates of stimuli to create conditions where participants are highly confident in their judgements about low visibility targets (see Sherman et al., 2015), or by varying decision boundaries orthogonally to stimulus strength (Bang & Fleming, 2018).

Our findings demonstrate that expectations shape precision estimates used at high levels of our cognitive architecture. We find these expectations alter confidence and awareness. However, Bayesian models suggest that precision estimation is important for diverse cognitive functions – including perception, learning and social cognition (Yon & Frith,2021). One possibility is that our cognitive system maintains a single representation of perceptual precision which is used to support all of these functions. However, we have suggested recently that different precision estimates are maintained at different levels of the hierarchy – and that expectations may exert a stronger influence on precision at higher levels (Yon & Frith, 2021). It is unclear from these findings whether expected precision will also change low-level perceptual inferences. It will thus be important for future work to establish whether expectations about precision exert a similar role on precision-weighted inferences in other domains.

For example, Bayesian models of multisensory integration suggest that observers combine signals from different modalities according to their estimated precision, lending more weight to more certain sensory channels (Alais & Burr, 2004; Ernst & Banks, 2002). Similarly, Bayesian models of prediction suggest that observers make perceptual inferences by combining incoming evidence with probabilistic expectations – leaning more on prior knowledge when the evidence is more ambiguous i.e., less precise (Olkkonen et al., 2014; Press et al., 2020; Yon, 2021). It is possible that the precision representations used to solve these combination problems are also shaped by expectations. For example, observers may learn to expect that their vision is unreliable in some contexts and use this *expectation* to control whether they rely on other senses or other kinds of knowledge when trying to make sense of the world around them. However, it also remains possible that these precision estimates are not shaped by expectations – and that low-level processes like perception use precision estimates that are more closely tied to the objective uncertainty of incoming signals rather than prior beliefs. Understanding whether and how expectations alter precision estimates at different levels of the cognitive hierarchy will constrain theorising about Bayesian models of the mind – clarifying when and whether beliefs about uncertainty detach from reality (Yon & Frith, 2021).

The current study provides strong support for influential Bayesian models of cognition, showing that agents combine incoming evidence with prior expectations to estimate the precision of their senses. These results begin to reveal the mechanisms we use to learn about uncertainty in our own minds and reveal that expectations about precision formed through such learning exert an influence on how we experience the sensory world and how much we trust our senses.

2.7. Supplementary Modelling

1. Weighted combination of point estimates

The model in Chapter 2 assumes that agents form an estimate (or inference) of the signal strength at timepoint *t* as a weighted combination of incoming evidence and prior expectation:

$$inference_t = w_{prior}(prior_t) + w_{evidence}(evidence_t) \qquad (1.1)$$

Where the weights on prior and evidence are defined as:

$$w_{prior} = 1 - w_{evidence} \qquad (1.2)$$

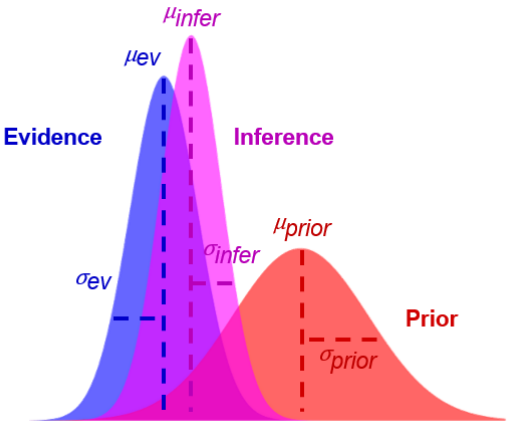
This model, in itself, makes no particular assumptions about how prior, evidence and inference are represented in the mind and brain – aside from assuming that agents have access to a point estimate of these variables at a given point in time. By trading only in point estimates, this weighted combination is similar to classic models of associative learning such as the Rescorla-Wagner learning rule (also known as the 'delta rule'; ((Dayan & Kakade, 2000; Rescorla & Wagner, 1972)). Indeed, substituting Eq 1.2 into Eq 1.1 and rearranging yields:

$$inference_t = prior_t + w_{evidence}(evidence_t - prior_t) \qquad (1.3)$$

This is identical to the Rescorla-Wagner rule, where a point estimate - $inference_t$ - is made by computing a prediction error - $(evidence_t - prior_t)$, - and using this difference value to update an initial expectation - $prior_t$. The size of this update is controlled by $w_{evidence}$, which is equivalent to the learning rate parameter in Rescorla-Wagner, usually denoted as α *(alpha)*. Given this equivalence and Eq 1.2, one could also think of w_{prior} is equivalent to $1 - \alpha$.

2. Relationship to models of Bayesian inference

However, while the model presented in Chapter 2 only assumes that agents form point estimates, it is also possible to connect these ideas to models of Bayesian inference. In particular, we could imagine that agents represent both the incoming evidence and their prior expectation as Gaussian distributions. These two Gaussian distributions are then combined together to form an inferred (posterior) distribution, which is also Gaussian. As shown in the figure below, each of these Gaussians is centred on a mean - μ *(mu)* - while the width of each distribution is controlled by the standard deviation - σ *(sigma)*.



The Bayes-optimal estimate for the inferred (posterior) distribution is given by combining the evidence and expectation distribution according to their precision, where precision is the inverse variance - $\frac{1}{\sigma^2}$. This means that more weight is given to the source of information (evidence or expectation) that is estimated to be most precise. In particular, the Bayes-optimal estimate for μ_{infer} is:

$$\mu_{infer} = w_{prior}\left(\mu_{prior}\right) + w_{ev}\left(\mu_{ev}\right) \tag{2.1}$$

Where:

$$w_{prior} = \frac{\frac{1}{\sigma^2_{prior}}}{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{ev}}} \tag{2.2}$$

And:

$$w_{ev} = \frac{\frac{1}{\sigma^2_{ev}}}{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{ev}}} \quad (2.3)$$

Because the denominators are the same, these weights sum to 1:

$$w_{prior} + w_{ev} = \frac{\frac{1}{\sigma^2_{prior}}}{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{ev}}} + \frac{\frac{1}{\sigma^2_{ev}}}{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{ev}}} = \frac{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{ev}}}{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{ev}}} = 1 \quad (2.4)$$

And so:

$$w_{prior} = 1 - w_{ev} \quad (2.5)$$

Note that Eqs 2.1 and 2.5 are identical to Eqs 1.1 and 1.2 which describe our model.

This means that it is possible to interpret the parameter w_{prior} in our model in Bayesian terms, as an agent's estimate of the precision (or confidence) of their expectations relative to the precision of the incoming evidence provided by the senses. Thus, if expectation and evidence are judged to be equally reliable $w_{prior} = .5$. In contrast, values of $w_{prior} < .5$ suggest that agents believe the incoming evidence is more reliable than prior beliefs (and vice versa if $w_{prior} > .5$).

However, while this equivalence between Bayesian inference and our model can be shown mathematically, in the present work we do not directly measure or manipulate σ^2_{ev} or σ^2_{prior}. As such, any value of w_{prior} used by an agent in our point-estimate model should be thought of as an *as if* Bayes-optimal inference (e.g., a participant whose behaviour is best fit by a value of $w_{prior} = .5$ is behaving in a Bayes-optimal fashion, if *they believe* incoming evidence and their prior beliefs are equally reliable, but we cannot verify whether this is true).

Future work directly measuring or manipulating variables implied by σ^2_{ev} or σ^2_{prior} (e.g., manipulating the uncertainty in evidence or expectations) will be important for determining whether we should conceptualise this kind of learning in fully Bayesian terms, rather than the simpler mechanics of the point estimate model described in Eqs 1.1 and 1.2.

Chapter 3: The neural mechanisms underpinning the formation and use of expectations about precision

3.1. Introduction

Chapter 2 provided support for the influential Bayesian model of metacognition which suggests that agents form probabilistic beliefs about how precise sensory information is *likely* to be and incorporate these expectations into perceptual precision estimates (Yon & Frith, 2021).

In this chapter, we probed the neural mechanisms underpinning the formation and use of these expectations about precision. For Experiment 5, we adjusted the task used in Experiment 3 for optimal use within the scanner environment and replicated the results. This acted as a behavioural pilot for the main 3T fMRI experiment (Experiment 6), investigating which brain areas track signal precision and how these representations are changed by expectations. In particular, we focused on areas previously implicated in the encoding of objective and subjective uncertainty (Geurts et al., 2022), including the dorsal anterior cingulate cortex (dACC), frontal pole and the insula – as well as visual brain regions like the middle temporal area (MT). To pre-empt our results, we find representations of sensory precision in the insula and MT, but only find these representations are modulated by expectations in the insula. Our results are consistent a possible 'prediction error' learning signal in the insula, explaining how higher-level brain regions may use our perceptual experiences to form and updated beliefs about precision.

3.2. Experiment 5 – Behavioural pilot

Experiment 5 acted as a behavioural pilot for the main fMRI experiment. We adapted the perceptual decision-making task from Experiment 3, where we asked participants to rate the clarity of moving dot clouds. Adaptations included the addition of coloured shape cues prior to the presentation of motion stimuli, the use of a circular, rather than linear, rating scale, and a change to the overall trial structure (see section '3.2.1. Procedure' for details). These adaptations optimised the paradigm for use in an MRI scanner environment and orthogonalised potential confounds for

fMRI decoding analyses (e.g., potential confounds between motion strength and dot colour, or between motion strength and the sensorimotor motor features of the response). The aim of this behavioural pilot was to investigate whether comparable results to Experiment 3 would be found even with these fMRI orientated adaptations.

3.2.1. Methods

Participants

Twenty participants (12 female, 8 male, mean age = 36.8, SD = 8.65) were recruited via Prolific. All participants were right-handed, reported normal or corrected vision and no history of psychiatric or neurological illness. All experiments were approved by the Research Ethics Committee at Birkbeck, University of London.

Participants were considered outliers if their clarity ratings were >2.5SDs away from the mean. Outliers were winsorized to values 2.5SDs away from the mean for inferential statistics. For Experiment 5, seven datapoints for three participants were winsorized. Data patterns and their statistical significance were unchanged by this adjustment.

Procedure

Participants completed an online perceptual decision-making task programmed in PsychoPy (Peirce et al., 2019) adapted from Experiment 3 in Chapter 2. Participants were presented with clouds of up- or downwards moving dots, which they rated the clarity of from '*completely clear*' to '*completely random*'. Each trial began with a fixation cross surrounded by a probabilistic shape cue (either a coloured square or triangle - 2000ms) before the presentation of a dot motion stimulus (750ms – see Figure 3.1A). In each cloud, the motion angle of each dot was drawn from a normal distribution centred on either 0 degrees (up) or 180 degrees (down). Varying the precision (or width) of this distribution makes it possible to create *clear* stimuli (where most dots follow the paths close to programmed 'mean' – up or down) or *ambiguous* stimuli (where motion directions are relatively more random, see Figure 3.1B).

After a blank screen (500ms), participants had up to twenty seconds to rate the clarity of the motion cloud using a circular scale, similar to that used by Geurts and colleagues (2022; see Figure 3.1C). Participants moved a red marker around the scale using their mouse. The scale was tapered at one end; placing the marker at the wider end of the scale would indicate participants thought the motion had been '*completely clear*', whereas markers placed towards the narrower end of the scale would indicate ratings of '*completely random*' motion. Participants were encouraged to use the full range of the scale. To prevent participant's clarity reports correlating with perceptual features of the scale, the orientation of the scale and the starting position of the response marker was randomised on each trial.

As in Experiment 3, probabilistic cues (here coloured shapes) allowed participants to form expectations about the clarity of motion signals on each trial, allowing us to investigate how such expectations bias perceptual judgements. For example, a green triangle could signal that motion clouds were likely to be ambiguous, while a blue square could signal that motion clouds would likely be clear. Unlike in Chapter 2, participants were explicitly informed about the nature of these cue mappings at the beginning of each block, and cue mappings were flipped halfway through the experiment to ensure that behavioural effects were not due to any particular mapping.

After 24 practice trials participants completed a main experiment of 384 trials. On 75% of trials observers experienced 'expected' events where the learned cues were valid (i.e., a blue square signalling clear motion was presented, and clear motion then followed) while on the remaining 25% stimuli were 'unexpected' (i.e., the blue square signalling coloured clear motion was followed by an objectively ambiguous motion). This manipulation of trial 'expectedness' allowed us to investigate how expectations about signal clarity influenced subjective clarity ratings.

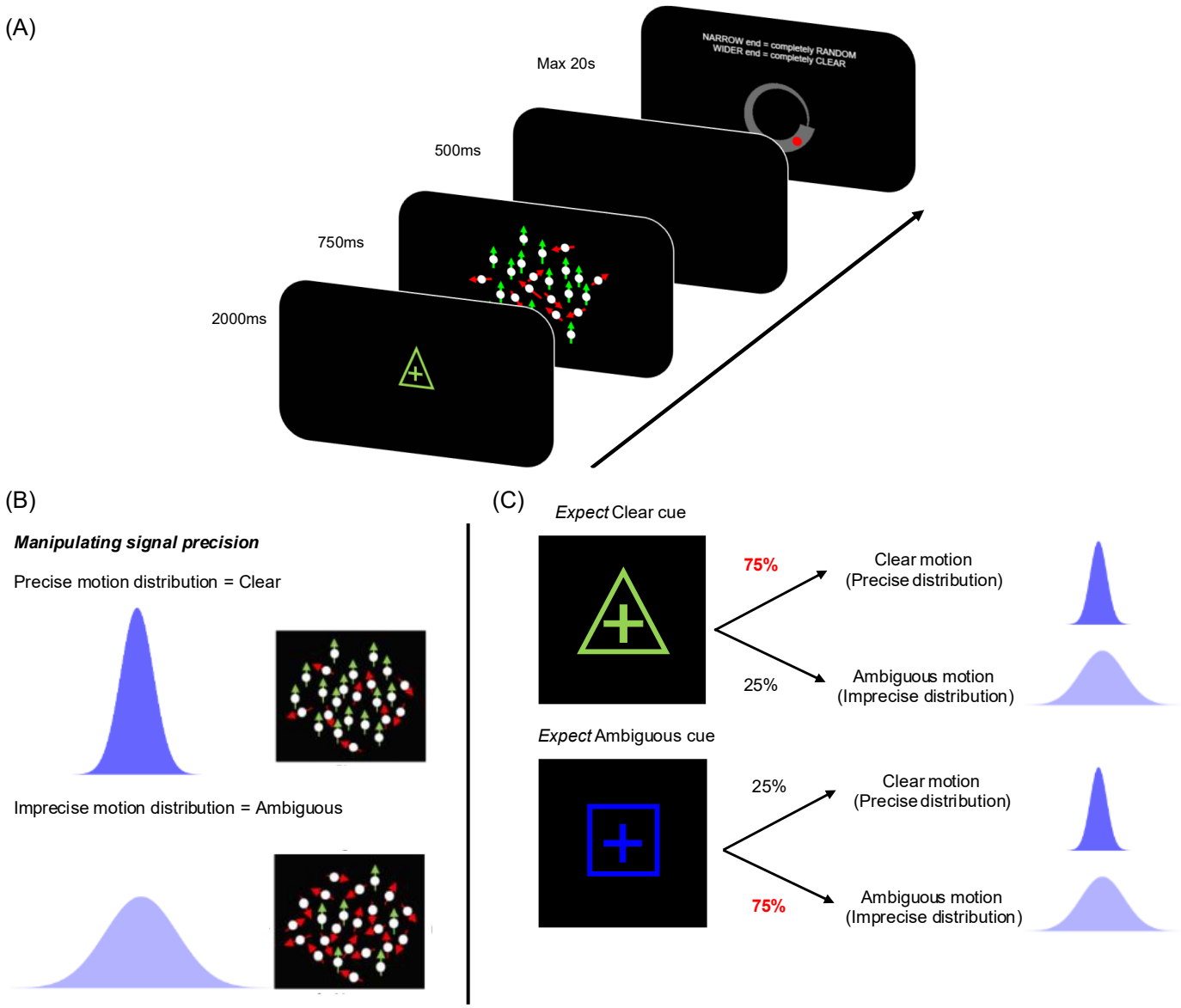


Figure 3.1 – Experimental task (A) Participants completed a motion perception task, rating the clarity of motion of brief motion clouds using a randomly-oriented circular scale. Coloured shape cues (green triangle, blue square) were used to manipulate expectations about the clarity of motion signals for each trial, e.g., if a green triangle and fixation cross was shown, the upcoming motion cloud was likely to display clear motion. **(B)** Motion clouds were manipulated to either be 'clear' or 'ambiguous' by drawing the dot motions from distributions with different 'precisions'. For 'clear' clouds, the majority of dots would move closely to 'mean direction' of the distribution (green arrows), while the remaining dots moved in more random directions (red arrows). For 'ambiguous' clouds, the majority of dots moved in more random directions away from the mean direction, making it difficult to decipher the overall motion direction. **(C)** During the 'test phase', participants would either experience 'expected' or 'unexpected' trials. For example, on an 'expected' trial, participants would see a 'expect clear' cue, followed by a 'clear' motion cloud. However, on a 'unexpected' trial, participants would see a 'expect clear' cue followed by an 'ambiguous' motion cloud.

3.2.2. Results

We investigated how actual and expected precision influence subjective clarity ratings. In particular, we were interested in whether expectations about the precision of the incoming visual signals would bias clarity ratings as in Experiment 3.

This was achieved by analysing clarity ratings with a 2 x 2 within-subjects ANOVA with factors of actual precision (*clear, ambiguous*) and expected precision (*expect clear, expect ambiguous*). Unsurprisingly, this analysis revealed a main effect of actual precision – with higher clarity ratings for stronger (mean = 0.732, SD = 0.08), than weaker stimuli (mean = 0.198, SD = 0.130; $F(1, 19) = 163.222$, $p<0.001$, $\eta_p^2 = 0.896$).

However, our key question concerns how *expected* precision alters subjective perceptions of clarity. Indeed, this analysis also revealed a main effect of expected precision – with higher clarity ratings on '*expect clear'* (mean = 0.972, SD = 0.106) compared to *'expect ambiguous'* trials (mean = 0.444, SD = 0.104; $F(1, 19) = 8.227$, $p = 0.010$, $\eta_p^2 = 0.302$, see Figure 3.2). There was not an interaction between these two factors.

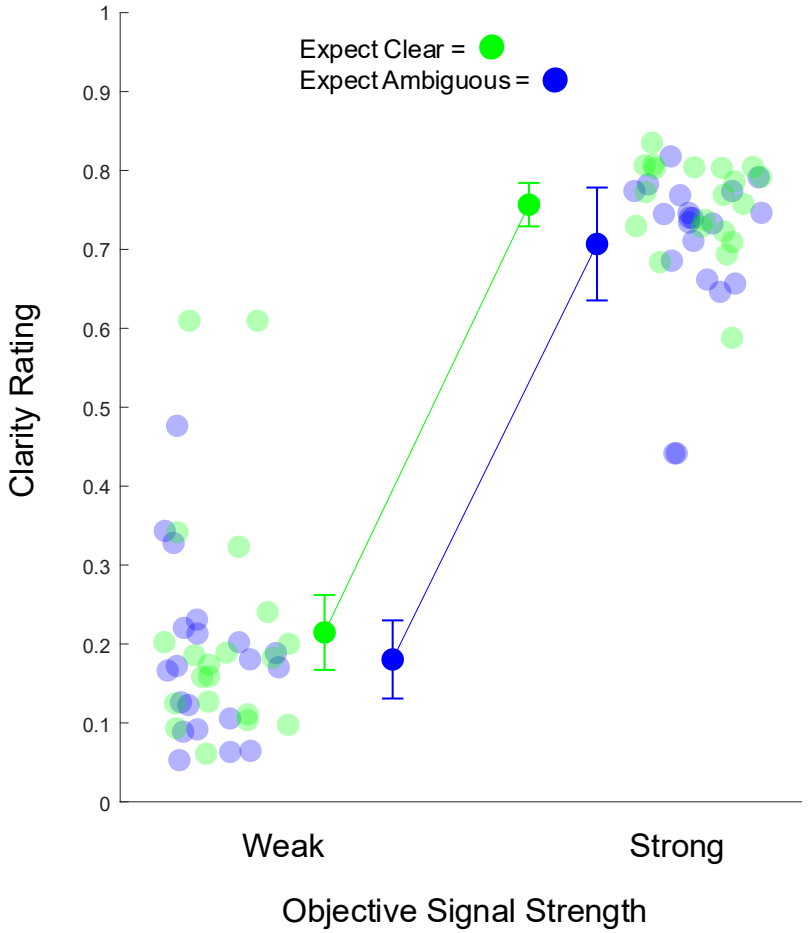


Figure 3.2 – Expected precision alters subjective clarity ratings: Participants reported significantly higher clarity ratings when they *expected* clearer signals regardless of the objective signal they received (i.e., the green line is above the blue). Error bars represent 95% confidence intervals on the mean.

3.2.3. Discussion

We found that observers were biased to give motion clouds higher clarity ratings when more precise signals were expected. This is consistent with findings from Experiment 3 and also suggests that participants continue to form and use expectations about perceptual precision in this revised experimental set up.

3.3. Experiment 6

In Experiment 6 we used functional magnetic resonance imaging (fMRI) to investigate the neural mechanisms responsible for tracking signal precision, and how these representations are modulated by expectations about signal precision. Here we used the same task as Experiment 5 with some adjustments for use in the MRI scanner (see section '3.3.1. Procedure' for details).

3.3.1. Methods

Participants

30 new participants (22 female, 8 male, mean age = 25.03, SD = 3.64) were recruited from local participant databases. The same inclusion criteria and outlier management procedures were used as in Experiment 5, with the addition that participants be 20-35 years old. This sample included one replacement for a participant who was not invited to the fMRI phase after the behavioural training session (see below), as their data suggested they could not discriminate between *clear* and *ambiguous* motion clouds (i.e., no difference in clarity ratings).

Procedure

Experiment 6 was split over two days. On the first day participants familiarised themselves with the behavioural task. On the second day they completed a brief refresher version of the task outside the scanner, before completing the same task in the MRI scanner while we recorded their brain activity.

Behavioural task

The task was almost identical to that used in the behavioural pilot (Experiment 5) with a few changes to event timings to optimise it for MRI analysis – namely, participants had a shorter response window (2 seconds) to register their clarity rating and a randomised jitter (0.5 – 5 seconds) was introduced between trials (see Figure 3.3 for more details). The refresher version of

the task, completed just before scanning, was a condensed version consisting of only 96 trials (75% 'expected' trials and 25% 'unexpected' trials).

In the scanner, participants completed eight runs of 48 trials each. As in the behavioural pilot, participants were exposed to 75% expected and 25% unexpected trials, and coloured shape cues signalling that participants should *expect strong* or *expect weak* signals were flipped mid-way (i.e., at the beginning or Run 5).

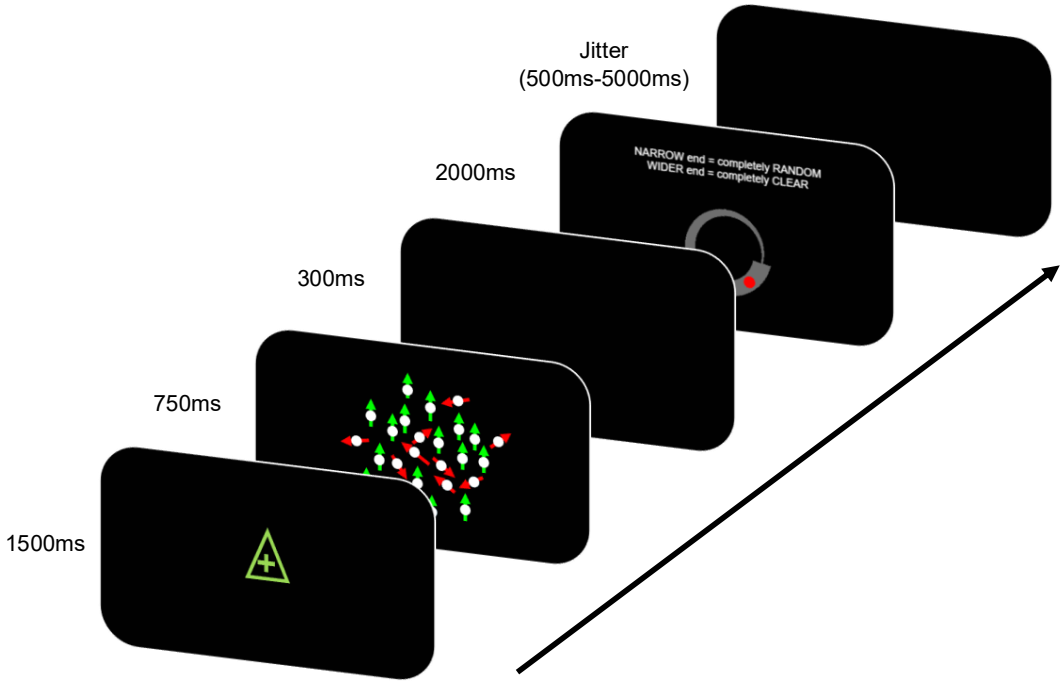


Figure 3.3 – Experimental task: We used the same task as Experiment 5, with small adjustments to response timing and inter-trial jitter to optimise fMRI data analysis.

fMRI acquisition and preprocessing

Images were acquired using a 3T (Prisma) MRI scanner (Siemens, Forchheim, Germany) and a 32-channel head coil. Functional images were acquired using an Echo Planar Imaging (EPI) sequence (ascending slice sequence, TR = 1.5s, TE = 35.02ms, 72 slices, voxel resolution 2mm isotropic). Structural images were acquired using magnetisation-prepared rapid gradient-echo (MP-RAGE) sequence (voxel resolution: 1mm isotropic).

Images were pre-processed in SPM12. The first eight volumes of each participant's data in each scanning run were discarded to allow for T1 equilibrium. All functional images were spatially realigned to the mean image and slice-time corrected. The participant's structural image was then co-registered to the mean functional scan and segmented to estimate forward and inverse deformation fields which can transform data from participant's native space into normalised space, and vice versa.

Defining regions of interest

We focused analyses on regions of interest (ROIs) previously implicated in the encoding of sensory precision and confidence judgements (see Figure 3.4). We used the same regions as Geurts and colleagues (2022) who found representations of objective and subjective precision in the dorsal anterior cingulate cortex (dACC) which is associated with environmental uncertainty (Behrens et al., 2007; Rushworth & Behrens, 2008), the frontal pole which is associated with subjective confidence (Fleming & Dolan, 2012), and the insula (which is associated with subjective uncertainty (Singer et al., 2009). We also added the middle temporal area (MT) as this area is implicated in the representation of visual motion (e.g. Born & Bradley, 2005). Bilateral masks of these ROIs were first created in normal space using the Neubert cingulate orbitofrontal connectivity-based parcellation (Sallet et al., 2013) for the dACC and the Harvard-Oxford cortical atlas (Desikan et al., 2006) for the frontal pole, MT and the insula. These masks were then transformed into each participant's native space using the inverse deformation fields created during preprocessing. These participant-specific ROI masks were used in all analyses, and all analyses were conducted in each participant's native space (i.e., functional data was not normalised).

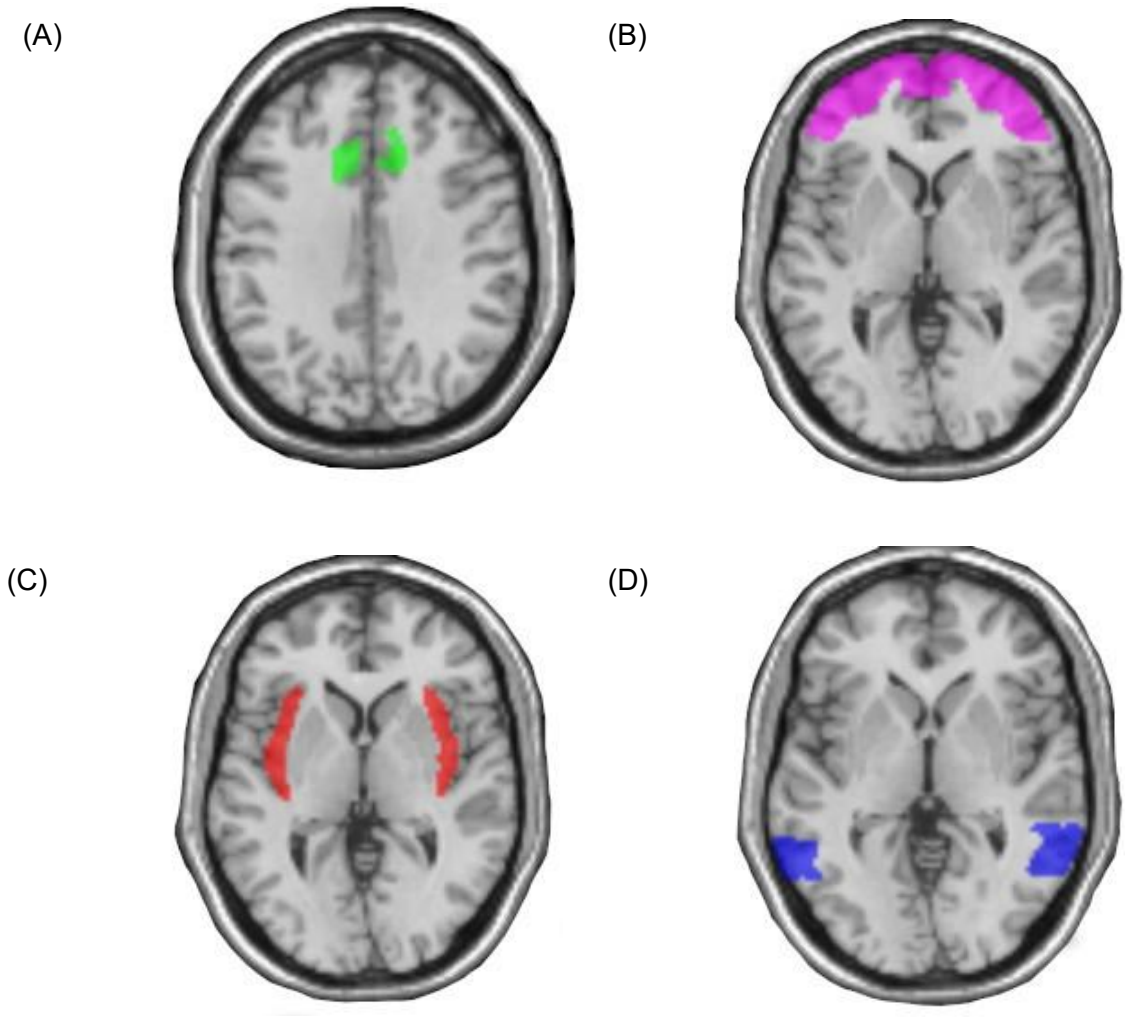


Figure 3.4 – ROIs for analyses: (A) dorsal anterior cingulate cortex (dACC) which is associated with environmental uncertainty, **(B)** Frontal pole which is associated with metacognition **(C)** the insula, which is associated with monitoring uncertainty and (**D)** the Middle Temporal area (MT), which is involved in representing visual motion.

3.3.2. Results

Behavioural analyses

First, we investigated whether expectations influenced subjective clarity ratings before and during scanning in the same manner as seen in Experiment 5. This analysis found the same effect of expected precision on clarity ratings was significant in the pre-scanning phase, with higher clarity ratings on *expect strong* (mean = 0.483, SD = 0.044) compared to *expect weak* trials (mean = 0.465, SD = 0.046; $F(1,26) = 27.525$, $p<0.001$, $\eta_p^2 = 8.256x10^{-4}$). The same effect was found

descriptively during scanning itself – (*expect strong* trials mean = 0.475, SD = 0.049; *expect weak* trials mean = 0.469, SD = 0.047*)* - but this did not reach significance (*F(1,26) = 2.449, p = 0.130,* η_p^2 *= 0.086)*, which may reflect somewhat noisier behaviour when participants complete the task inside the scanner. However, these results broadly suggest that observers were able to track and use expectations about precision and integrate these into their inferences about signal clarity.

fMRI analyses

We conducted multivariate decoding analyses to investigate whether each of our ROIs contained representations of sensory uncertainty, and whether these representations were altered by expectations about precision. All analyses were conducted using the TDT toolbox in MATLAB.

First, to establish whether each ROI contained information about sensory uncertainty we constructed a classifier (support vector machine) which was trained to discriminate objectively strong trials from objectively weak trials, collapsed over expectations. This achieved initially by specifying a general linear model (GLM) in SPM12 which included event-related (stick) regressors locked to the onset of each 'weak' or 'strong' motion cloud separately in each run, alongside nuisance regressors capturing participant head movement, all of which were convolved with the canonical haemodynamic response function. Regressing this model against BOLD activity yields a total of 16 beta images of interest (one 'strong' and one 'weak' image, for each of the 8 scanning runs). These beta images are then used as training and test data for a linear SVM which learns to discriminate between strong and weak trials, using a leave-two-out cross-validation procedure. Crucially, this procedure 'leaves out' one run from each half of the experiment (Runs 1-4 and Runs 5-8) to orthogonalise stimulus strength and the coloured shape cues. The classifier's accuracy was calculated as the proportion of correctly classified images during the decoding steps.

This analysis revealed that decoding of stimulus strength was above chance in the insula (mean decoding = 53.5%, SD = 7.592, *t(29) = 2.556, p = 0.016, dz = 0.467)* and area MT *(*mean decoding = 54.9%, SD = 9.242, *t(29) = 2.906, p = 0.007, dz = 0.530)* suggesting that were was reliable information about objective signal strength (i.e., the classifier could discriminate between 'strong'

and ‘weak’ signals). In contrast, decoding of objective sensory uncertainty was not significantly above chance in the dACC *(*mean decoding = 52.9%, SD = 8.967, *t(29) = 1.787, p = 0.084, dz = 0.326)* or frontal pole *(*mean decoding = 51.640%, SD = 8.969, *t(29) = 1.001, p = 0.325, dz = 0.183)* did not (see Figure 3.5).

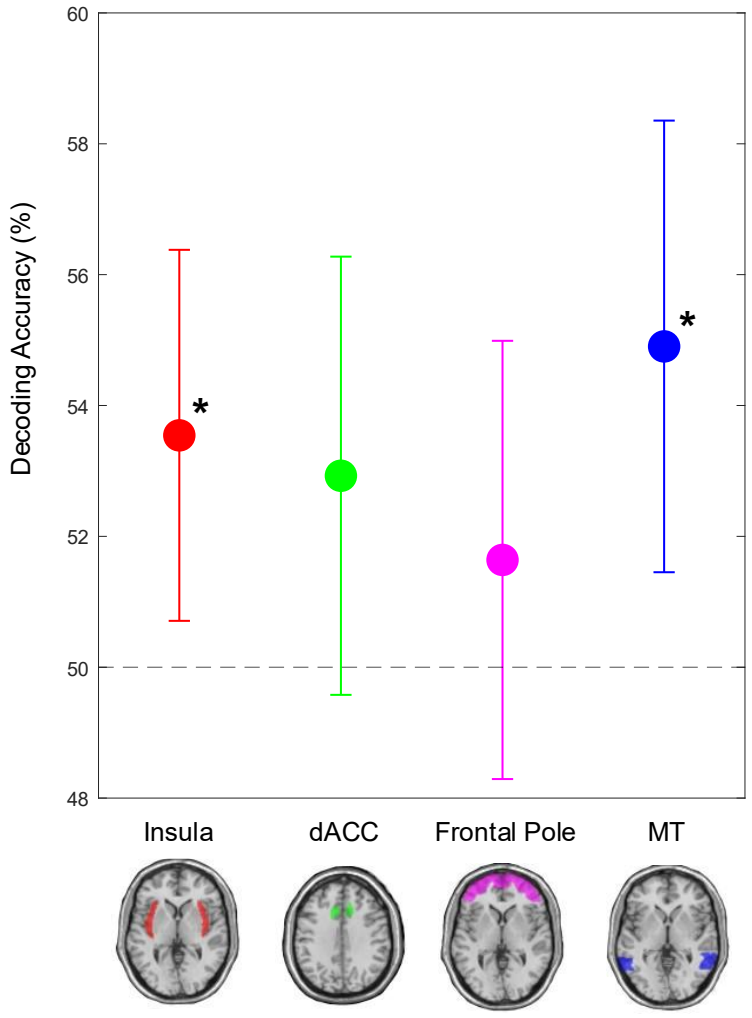


Figure 3.5– Insula and area MT show superior decoding accuracy of objective signal strength: in comparison to the dACC and FP, the insula and area MT show significantly higher decoding accuracy, indicating that the classifier was able to reliably discriminate between objectively strong and objectively weak trials in these areas. Error bars represent 95% confidence interval on the mean.

In the second phase of our analyses, we then investigated whether the representations of sensory uncertainty we identified in our first step were manipulated by expectations. This involved constructing another GLM with separate event-related regressors for *expected* and *unexpected* strong and weak stimuli (e.g., *expected strong* clouds, *expected weak* clouds, *unexpected strong* clouds and *unexpected weak* clouds). Given the fact that there were (by design) three times more expected than unexpected trials, to keep the signal-to-noise ratio consistent between conditions (and following Kok et al., 2012) we randomly divided expected events into three separate regressors – so that all expected and unexpected beta images were estimated from the same number of trials.

With this GLM in hand, we then conducted equivalent decoding analyses discriminating 'strong' from 'weak' trials – but separately for *expected* and *unexpected* conditions. The logic of this analysis is that if representations of sensory uncertainty are altered by expectations about precision, decoding accuracy should differ between expected and unexpected conditions. Separate *expected* and *unexpected* classifiers were trained and tested again using the same leave-two-out cross-validation procedure. Decoding accuracy was again evaluated by averaging the proportion of correct classifications over decoding steps, and overall *expected* accuracy was estimated by averaging together the results from three randomly-divided *expected* classifiers. In the final step of the analysis, we compared decoding accuracy between *expected* and *unexpected* conditions in each of our ROIs using a paired sample t-test.

These results revealed that there was superior decoding in the insula for '*Unexpected'* trials (mean decoding = 54.3%, SD = 10.438) in comparison to *'Expected'* trials (mean decoding = 49.9%, SD = 6.354, $t(29) = -2.300$, $p = 0.029$, $dz = -0.420$). However, there was no significant difference in decoding accuracy between trial types in area MT ($t(29) = 0.002$, $p = 0.998$, $dz = 3.982 \times 10^{-4}$, see Figure 3.6).

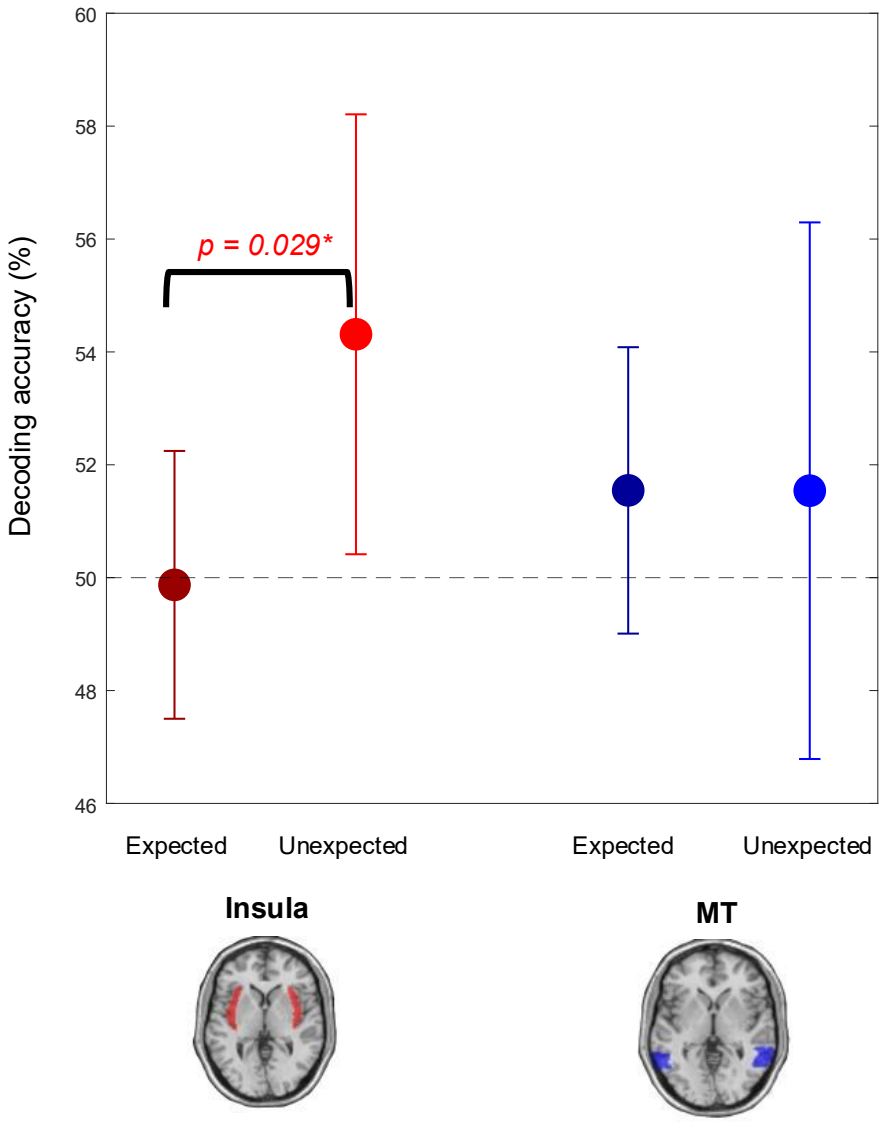


Figure 3.6 – Superior decoding for *unexpected* trials in the insula: Results showed a significant difference in decoding accuracy between trial types in the insula. We found no difference in decoding accuracy between *expected* and *unexpected* trials in area MT. **Values below 50 on this graph indicate decoding accuracies below chance (50%).** Error bars represent 95% confidence on the mean.

3.3.3. Discussion

Experiment 6 found representations in the insula and area MT tracking of the sensory uncertainty of incoming signals. More importantly, only the insula showed modulation of these representations according to expectations. In particular, there was superior decoding of sensory precision on '*Unexpected*' trials. One possibility is that this pattern arises because the insula encodes 'precision

prediction errors' that capture the difference between the precision we are expecting and the precision we receive. These prediction error signals could be important for forming and updating expectations about signal strength, in the same way that prediction errors are thought to drive learning in other sensory and reward domains.

3.4. General discussion

In previous experiments we found that agents combine incoming evidence with learned expectations to estimate the precision of sensory information. However, it remained unclear how information about sensory precision is encoded in the brain and how representations in these areas are modulated by participant's expectations about signal precision. The results of this chapter identify the insula as a potentially pivotal site in the brain where sensory precision is encoded, where representations are modulated by expectations about precision. In particular, we found superior decoding for stimuli with unexpected precision in comparison to expected events. These results could be interpreted as a form of 'precision prediction error' which could play an important role in forming and updating expectations about signal strength.

Given that our behavioural results have previously indicated a bias towards expectations (i.e., when people expect more precise sensory signals, they give higher clarity ratings), one may have predicted that neural decoding results would have followed the same pattern (i.e., superior decoding for *expected* events). However, we find the complete opposite, with superior decoding accuracy for *unexpected* events – a possible 'precision prediction error' effect. Although our behavioural and neural results seem to be at odds with each other, with a bias in different directions, it is not inconceivable that results we pertain through neural decoding methods would manifest differently at a behavioural level.

Our neural results could point to the possibility that the insula plays a role in encoding 'precision prediction errors' that capture the difference between precision we are expecting and the precision we receive, the disparity between expectation and reality. Such prediction errors could enable agents to represent information with higher fidelity, with higher quality representations of signals when they are surprising, allowing agents to update their prior expectations. This could be crucial

in the formation of expectations about precision that we see in Chapter 2 and the behavioural results earlier in this chapter. Similar kinds of ‘awareness prediction errors’ are described in models of subjective consciousness such as Fleming's (2020) ‘higher-order space state’ model. However, this is the first empirical demonstration of how and where such ‘precision prediction errors’ could exist.

Researchers have already suggested that it would be difficult for the brain to monitor sensory uncertainty based purely on incoming signals alone, and observers may thus combine incoming signals with expectations to form precision estimates (Friston, 2018; Yon & Frith, 2021). Our results bring credence to this claim, showing that information about sensory uncertainty is monitored in the insula and modulated by participant’s expectations about sensory precision.

We focused our analyses on regions previously implicated in the ‘common coding’ of objective sensory uncertainty and subjective confidence – including the insula, alongside other prefrontal and cingulate regions like the frontal pole and dACC. It is noteworthy that in our studies we did not find representations of sensory uncertainty in prefrontal or cingulate regions – contrary to prior work (e.g., Geurts et al, 2022). One possibility for this inconsistency is that – in our study – participants were not actually required to construct confidence reports, and separate neural mechanisms may be involved in the representation of decision confidence and subjective visibility (Bang & Fleming, 2018, Mazor et al, 2022). For example, elegant neuroimaging work by Bang and Fleming (2018) and Mazor and colleagues (2022) has found neural representations for the encoding of vividness of sensory signals, which are distinct from those encoding decision confidence. Our analyses support these results. We only found above chance decoding of signal precision in the insula and area MT but no such decoding in the frontal pole and dACC. This may suggest that some apparent representations of perceptual precision in fact reflect representations that are only deployed when internal signals (like precision estimates) are translated into representations of decision confidence. One possible way to test this would be incorporate explicit decisions and confidence ratings into our paradigm – as if this explanation is correct, representations of confidence/uncertainty in areas like the frontal pole may then emerge (Fleming & Dolan, 2012).

Despite finding representations of signal precision in both MT and the insula, we only see that expectations modulate representations in the insula and *not* area MT. This suggests that expectations are formed and integrated into precision representations at higher-level processing stages rather than lower-level perceptual circuits. This is consistent with our empirical findings (e.g., in Chapter 2) that expectations about precision influence 'high level' features of perceptual decisions (like subjective confidence) but leave 'low level' features (like perceptual sensitivity) unchanged, and is also consistent with theoretical models which suggest expectations about precision should be more influential as we progress up the hierarchy of the mind and brain (Yon & Frith, 2021).

Experiments 5 and 6 provide further support for influential Bayesian models of cognition, showing that agents combine incoming evidence with prior expectation to estimate the precision of their senses (Experiment 5), and revealing a possible locus of these predictive computations in the insula (Experiment 6). These results bring us one step closer to understanding the mechanisms involved in how to perceive the sensory world.

Chapter 4: The role of expected precision in audition and anomalous perception

4.1. Introduction

In previous chapters, we have provided extensive evidence of agents monitoring, and forming expectations about the precision of visual stimuli (Experiments 1-5) and have also identified the neural mechanisms involved in such processes (Experiment 6). However, up until this point, we have focused solely on the visual domain. There are established Bayesian models of multisensory integration which suggest that observers combine signals from different modalities according to their estimates of precision, lending more weight to more certain sensory channels (Alais & Burr, 2004; Ernst & Banks, 2002). These models lean heavily on the assumption that agents track signal

precision and form expectations about such information across all the sensory domains, however, this is yet to be explored (Yon & Frith, 2021).

Understanding the nuances of how expectations alter precision estimates across the different sensory domains is particularly important given theoretical accounts which use this idea to explain unusual experiences and atypical cognition. For example, a prominent explanation of hallucinations in psychosis suggests that these unusual experiences arise because patients hold inappropriate beliefs about the relative precision of incoming sensory signals and top-down predictions, leading to a disproportionately strong weight on prior expectations when perceiving the world (Corlett et al., 2019). These accounts depend on the idea that beliefs about precision can be false, and this could arise if observers hold the wrong expectations about precision. However, theories that explain atypical cognition by appealing to atypical precision weights rely on the assumption that precision weights can indeed be learned. We have shown that this is possible across Experiments 1 to 6, but only in the visual domains. The formation of expectations about precision is yet to be tested across other sensory domains, such as audition (Yon & Frith, 2021) – which may be particularly important given that hallucinations in the auditory domain are particularly prevalent (Corlett et al., 2019).

In this chapter we explored whether 'expected precision' mechanisms identified in previous chapters can be generalised to the auditory domain, specifically to our perception of speech. We also explored whether disruptions to such mechanisms could explain unusual experiences such as hallucinations. Across two experiments we adapted the visual decision-making task used in previous chapters to probe 'expected precision' in the auditory domain. Participants then also completed the Cardiff Anomalous Perceptions Scale (CAPS) to measure their proneness to hallucination-type experiences, allowing us to connect these cognitive processes to hallucination-like phenomenology participants may experience.

To pre-empt our results, in Experiments 7 and 8 we find comparable results to those seen in our visual experiments; when participants *expected* clear auditory signals, they reported higher clarity ratings for the speech stimuli. This suggests that listeners integrate expectations about the reliability or *precision* of upcoming auditory stimuli into ratings of subjective clarity. When

examining the relationship between participant's reliance on their precision expectations and their tendency to display hallucination-type experiences, we found that there was a negative correlation between the two factors (but this only reached significance in Experiment 7). This reflected that those who are more prone to hallucination type experiences displayed a smaller Expectation Effect, or less of a bias towards expected clarity. This hints at a potential impairment in the formation and use of precision expectations in those who are more prone to hallucination-like experiences.

4.2. Experiment 7

Experiment 7 investigated whether listeners form and use expectations about the precision of incoming auditory signals to make judgements about the reliability of perceptual information.

Participants completed a perceptual decision-making task, rating the clarity of speech stimuli as *'clear'* or *'ambiguous'*. As in previous chapters, probabilistic cues signalled whether sensory signals would likely be strong or weak. We investigated how expectations established by these cues biased the subjective clarity of the speech stimuli and whether effects of these expectations were related to participants individual hallucination-proneness.

4.2.1. Methods

Participants

Fifty-one participants (37 female, 12 male, 2 other, mean age = 31.02, SD = 9.197) completed Experiment 7. All participants reported normal or corrected vision and hearing, with no history of psychiatric or neurological illness. Participants took part as part of a research methods lab class on the MSc Psychology programme. The sample size was determined by the number of students in the class who completed the online study within a 24-hour window. Some participants did attempt to take part but were not included in the final sample because datasets were incomplete (five) or corrupted (five). This experiment was approved by the Research Ethics Committee at Birkbeck, University of London.

For the clarity rating data, participants were considered outliers if their ratings were >2.5SDs away from the mean. Outliers were winsorized to values 2.5SDs away from the mean for inferential statistics. For Experiment 7, only 8 datapoints for 3 participants were winsorized. Data patterns and their statistical significance were unchanged by this adjustment.

For the CAPS data, Mahalanobis distance values were calculated for each participant's overall CAPS score. Cases with Mahalanobis distance values exceeding the critical chi-square value ($p<.001$) were considered outliers and removed from further analysis, leaving a sample of 50 participants (36 female, 12 male, 2 other, mean age = 31, SD = 9.289). Data patterns and their statistical significance were unchanged by this outlier removal.

Speech Stimuli

During the task, participants were played one of four speech stimuli of either the word *'Pie'* or *'Tie'* (recorded in a male, Southern British accent). Original recordings of these words were sourced from Sohoglu and Davis (2020).

The clarity of these speech stimuli was manipulated using noise-vocoding (Shannon et al., 1995). We used a similar noise-vocoding procedure as that used by Zoefel and colleagues (2020). Firstly, each word was first filtered into 16 logarithmically spaced frequency bands and the amplitude envelopes extracted for each band. The envelope of these frequency bands was then mixed with the broadband envelope of the original speech signal at proportions of 0% to 100% to give an envelope for each frequency band, used to modulate the noise in the respective frequency. The resulting signals were then re-combined to yield 100 16-/1-channel vocoded speech mixes. Speech stimuli with higher vocoded values had higher precision (i.e., it was easy to identify the spoken word), whereas stimuli with lower vocoded values had lower precision (i.e., the speech becomes progressively unintelligible). For full details on the vocoding process please see Zoefel and colleagues (2020).

From these 100 vocoded versions of *'Pie'* and *'Tie'*, we identified a *'clear'* stimulus (i.e., a recording with a high vocoded value, 92% morph) and an *'ambiguous'* stimulus (i.e., a recording with a low

vocoded value, 66% morph) were selected for use in the main experiment. All speech stimuli were 1000ms in length.

Procedure

Experiment 7 was modelled as an auditory version of Experiment 3. Participants completed an online perceptual decision-making task programmed in PsychoPy (Peirce et al., 2019), rating the clarity of speech stimuli (Figure 4.1A). Each trial began with a silent fixation period (500ms), during which a white speaker symbol would be displayed on screen. This speaker symbol would alternate between white (silent periods) and grey (when a speech stimulus was being played). One of the four speech stimuli would then be played (*clear pie, clear tie, ambiguous pie* or *ambiguous tie,* 1000ms). After another silent fixation period (500ms), participants then gave a clarity rating on a continuous scale ranging from *'ambiguous'* to *'clear'* (Figure 4.1B).

Critically, probabilistic cues (a single beep or double beep – see Figure 4.1C) allowed listeners to form expectations about the clarity of the auditory signals on each trial. For example, for a given listener, when a single 'beep' was played the following speech stimuli was likely to be of low clarity (i.e., difficult to discern whether there was a word within the auditory stimuli). In contrast, when a double 'beep' was played, speech stimuli were likely to have high clarity (i.e., easy to discern that there was a word within the auditory stimuli). This allowed us to investigate how expectations about auditory precision biased clarity ratings. Beep mappings were counterbalanced across participants.

The experiment comprised of 480 trials. The first 160 trials acted as an initial training phase to establish expectations about the beep cues. In this training phase, 90% of trials were 'expected' (e.g., listeners were *expecting* to hear a *clear* speech stimulus and were played a *clear* stimulus), and 10% were 'unexpected' (e.g., listeners were *expecting* a *clear* speech stimulus but were actually played an *ambiguous* one). Participants then completed a 320-trial test phase, where the ratio of expected/unexpected trials was adjusted to 75/25%.

Finally, participants completed the Cardiff Anomalous Perceptions Scale (CAPS) questionnaire (Bell et al., 2006), which is a validated measure of perceptual anomalies, and used here as a

measure of hallucinations proneness. A high score on the CAPS would indicate an individual who experiences a high level of hallucination-type experiences.

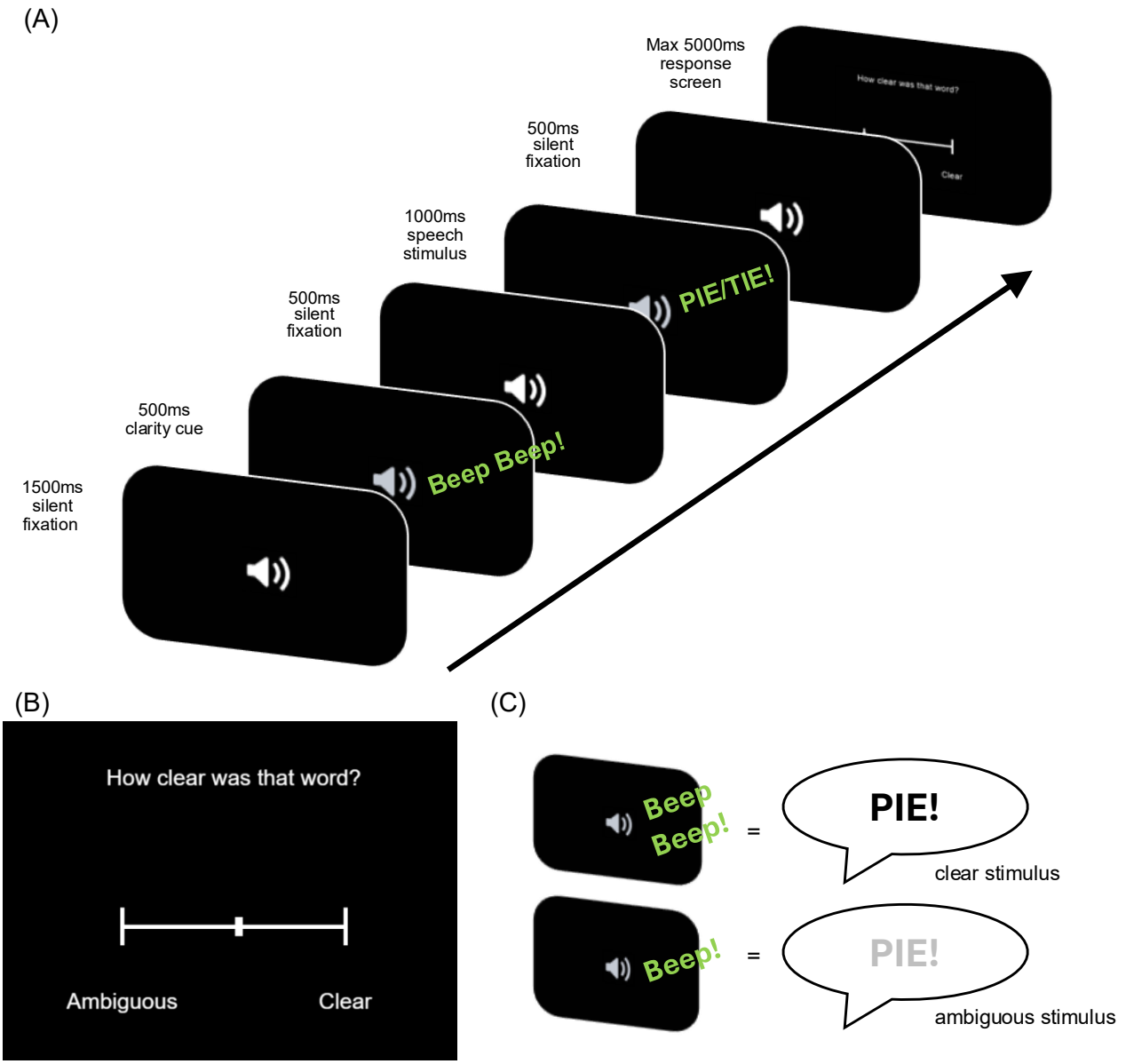


Figure 4.1 – Experimental task: (A) Participants completed an auditory perception task, judging the clarity of speech stimuli that could vary in their actual and expected ambiguity. **(B)** Participants moved a marker across this scale using their mouse, clicking to give their final rating of the clarity of the speech stimuli **(C)** Probabilistic beep sequences manipulated expectations about the strength (or clarity) of speech stimuli for each trial, e.g., if a participant heard a double beep sequence they could expect upcoming speech signals to be clear.

4.2.2. Results

We investigated how actual and expected precision altered perception of speech stimuli by analysing subjective clarity ratings - taken as a measure of the subjective precision of the speech stimuli. In particular, we were interested in whether expectations about the precision of the incoming auditory signals would bias clarity ratings. We also investigated whether hallucination-proneness was connected to how far listeners relied on their expectations when making these clarity judgements.

To investigate this first question, we analysed clarity ratings across the test phase of the experiment using a 2 x 2 within-subjects ANOVA with factors of actual precision (clear, ambiguous) and expected precision (expect clear, expect ambiguous). Unsurprisingly, this analysis revealed a main effect of true auditory precision, with higher clarity ratings for clear speech stimuli (mean = 0.649, SD = 0.139), than ambiguous speech stimuli (mean = 0.418, SD = 0.191; $F(1, 50) = 81.254$, $p<.001$, $\eta_p^2 = 0.619$). Critically, this analysis also revealed a main effect of expected precision, such that participants gave higher clarity ratings when they expected clearer signals (mean = 0.542, SD = 0.165) compared to when they expected ambiguous ones (mean = 0.525, SD = 0.165; $F(1, 50) = 9.480$, $p = 0.003$, $\eta_p^2 = 0.159$). These factors did/didn't interact ($F(1,50) = 0.197$, $p = 0.659$, $\eta_p^2 = 0.004$).

This suggests that participants did indeed form expectations about the reliability or *precision* of the upcoming auditory stimuli, and these expectations were integrated into subjective perceptions of clarity (see Figure 4.2).

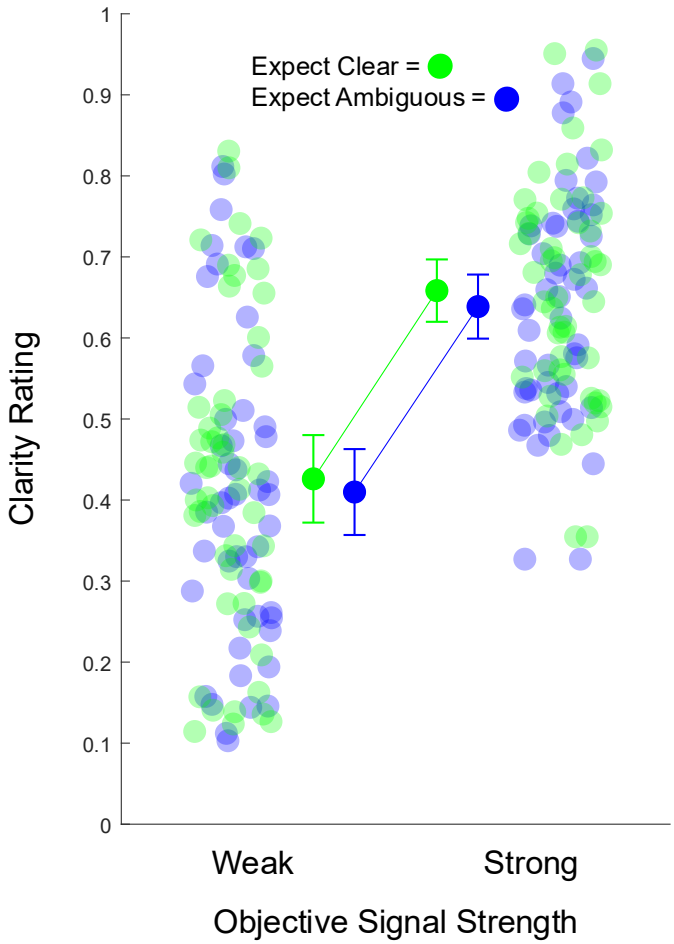


Figure 4.2 – Expected precision alters subjective clarity ratings: Participants reported significantly higher clarity ratings when they *expected* stronger speech stimuli ($p = 0.003$). Solid dots with error bars represent group level means with 95% confidence intervals on the mean. Translucent dots show individual subjects mean clarity ratings for different trial types.

To assess the extent to which people rely on their expectations to make judgements on the clarity of the auditory stimuli and whether this is related to participant's hallucination-proneness, we first calculated individual's 'Expectation Effect'. This was done by calculating the difference between average clarity ratings on trials where listeners *expected* strong signals and trials where listeners *expected* weak signals. A higher Expectation Effect score would indicate that a listener is relying on their expectations about precision. A lower Expectation Effect score would suggest similar perceptual ratings irrespective of expectations, indicating that a listener is relying less on expectations about which signals are likely to be ambiguous or reliable.

We found that there was a significant negative correlation between CAPS score and Expectation Effect (*ρ (48) = -0.318, p = 0.024)*, suggesting that those with a higher CAPS score (and therefore more prone to hallucination type experiences) displayed a smaller Expectation Effect (Figure 4.3).

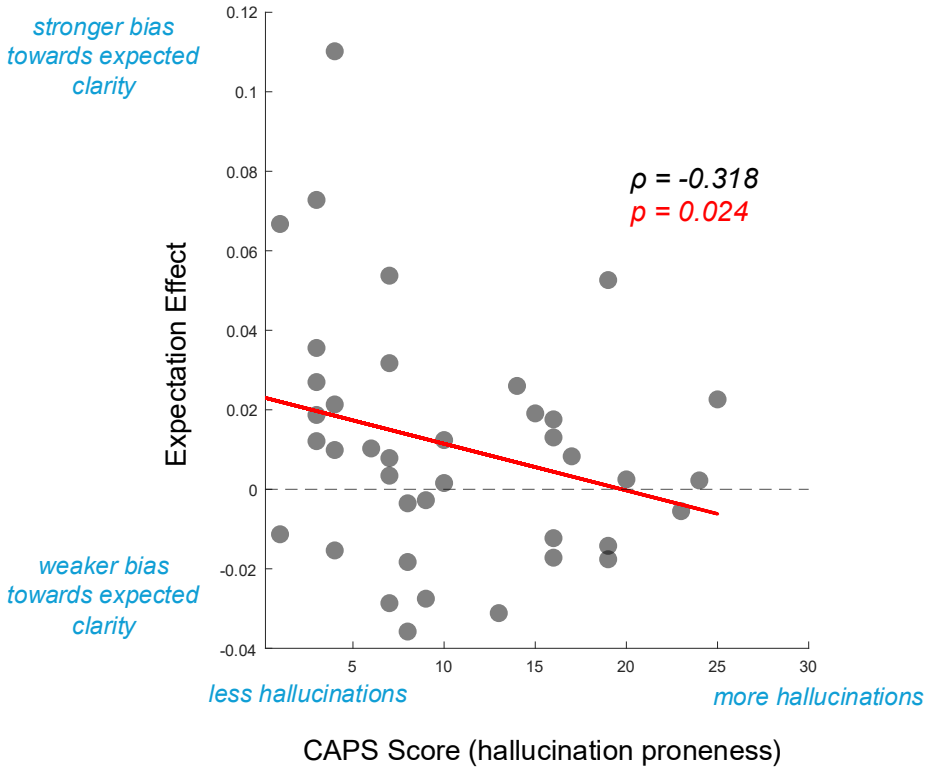


Figure 4.3 - Significant negative correlation between CAPS score and Expectation effect: This suggests that those participants who are more prone to hallucination-type experiences show a lower expectations effect.

4.2.3. Discussion

Here we find a similar pattern of results as Experiments 3-6; when listeners *expected* clearer speech stimuli, they gave higher clarity ratings than when they *expected* more ambiguous speech sounds. This again provides evidence consistent with the Bayesian idea that expectations about signal strength biases subjective awareness which is generalisable across visual and auditory domains. We also found a significant negative correlation between CAPS score (hallucination

proneness) and Expectation Effect – those more prone to hallucinations showed a weaker Expectation Effect, suggesting a potentially reduced ability to incorporate expectations about precision into judgements about the clarity and reliability of incoming sensory signals.

4.3. Experiment 8

Experiment 8 was exactly the same as Experient 7 but conducted with a larger sample size to investigate whether results from Experiment 7 were replicable.

4.3.1. Methods

Participants

A new sample of 194 participants (100 female, 91 male, 3 other, mean age = 32.32, SD = 7.98) were recruited via Prolific. This sample size was selected to provide 80% power to detect at least the same effect size as Experiment 7 ($\eta_p^2 = 0.159$). The same exclusion and outlier identification procedures were used as in Experiment 7. This sample included replacements for eight participants who were excluded for failing to complete at least 90% of trials across the training and test phase. An identical approach was used for identifying and managing outliers as in Experiment 7. No adjustments were applied in the analysing of subjective clarity ratings across conditions, but four extreme outliers were removed from the correlational analysis investigating connections between expectation effects and hallucination-proneness. No adjustments changed the significance of statistical patterns observed.

Procedure and Paradigm

The stimuli, paradigm and procedure were exactly the same as that used in Experiment 7.

4.3.2. Results

We carried out exactly the same analysis as Experiment 7. We first compared the subjective clarity ratings on trials where auditory signals were objectively stronger or weaker. Again, clarity ratings were higher when signals were objectively stronger (mean = 0.683, SD = 0.15), than when they were objectively weaker (mean = 0.428, SD = 0.165; *$F(1, 193) = 468.807$, $p<0.001$, $\eta_p^2 = 0.708$).*

Consistent with Experiment 7 we found that '*expect clear'* trials elicited higher clarity ratings (mean = 0.569, SD = 0.154), than *'expect ambiguous'* trials (mean = 0.542, SD = 0.162; *$F(1, 193) = 16.501$, $p<0.001$, $\eta_p^2 = 0.079$*).

These results replicate Experiment 7; suggesting that listeners integrate expectations about the reliability or *precision* of upcoming auditory stimuli into ratings of subjective clarity (Figure 4.4).

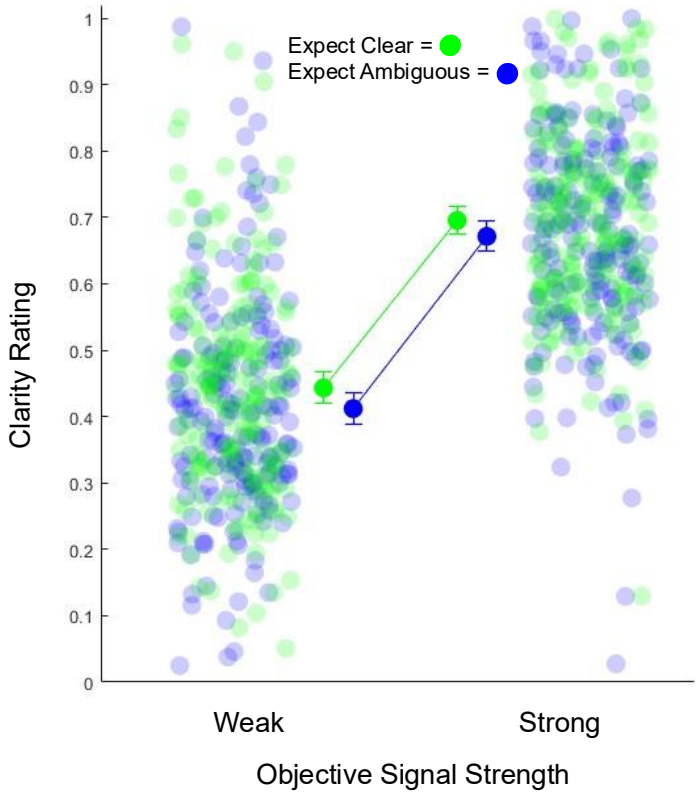


Figure 4.4 - Expected precision alters subjective clarity ratings: Similarly to Experiment 7, participants reported significantly higher clarity ratings when they *expected* stronger speech stimuli (*$p<0.001$*). Solid dots with error bars represent group level means with 95% confidence intervals on the mean. Transparent dots show individual subjects mean clarity ratings for different trial types.

We again assessed whether there was a relationship between participant's 'Expectation Effect' and CAPS score. As in Experiment 7, there was a numerically negative correlation between CAPS score and Expectation Effect (see Figure 4.5), but this was not significant (*ρ (188) = -0.018, p = 0.807).*

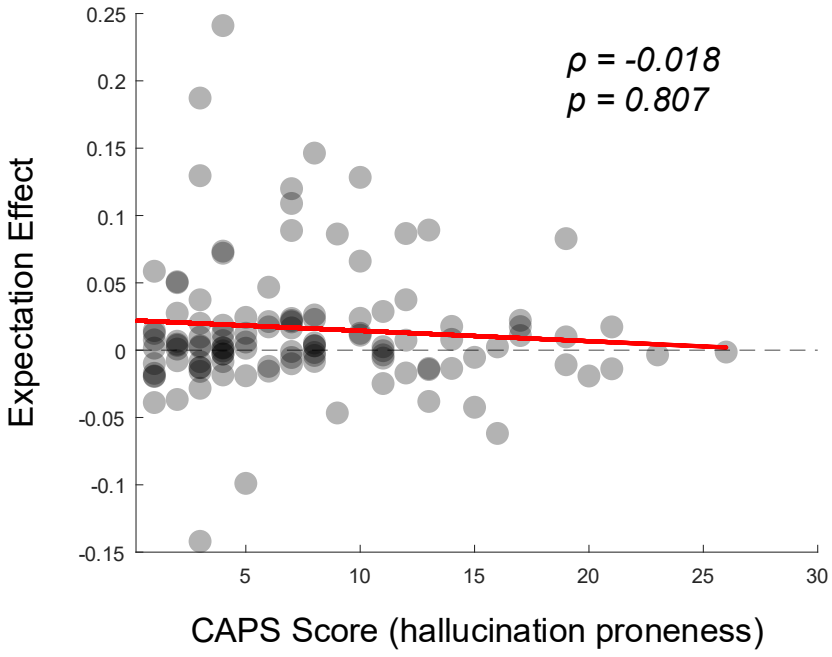


Figure 4.5 – No relationship between CAPS score and Expectation effect: as participant CAPS score increases, Expectation effect decreases, however, this relationship is not significant.

4.3.3. Discussion

In terms of our main task-based effects, results from Experiment 8 replicated those of Experiment 7; listeners reported higher clarity ratings when they *expected* clearer auditory signals, in comparison to when they expected more ambiguous auditory signals. When looking at the relationship between CAPS score and Expectation Effect, although we saw the same descriptively negative association as seen in Experiment 7 (i.e., those with a higher CAPS score displayed a lower Expectation Effect), this correlation did not reach significance.

4.4. General discussion

In this chapter, we explored whether 'expected precision' mechanisms identified in previous chapters could be generalised from vision to speech perception and whether these mechanisms could possibly explain atypical cognition such as hallucinations. Across two experiments, we found that it is indeed possible for agents to form and use precision expectations about incoming speech stimuli i.e., when they *expected* clearer speech, they were more likely to give higher clarity ratings than when they *expected* ambiguous speech. These results support the idea that agents combine incoming evidence with learned expectations to estimate the precision of sensory information, not just in the visual domain but also in the auditory domain.

Furthermore, in Experiment 7 we found a significant negative correlation between participant's CAPS score and their Expectation Effect. Experiment 8 resembled this pattern of results; however, the pattern did not reach significance. This suggests that the fundamental cognitive and brain mechanisms identified in Chapters 2 and 3 could be plausibly disrupted in those who experience atypical cognition such as hallucinations – though more investigations are needed to test the robustness of this possible association. For example, future work could explore this relationship further by recruiting a more diverse population continuum of CAPS scores. This could be achieved by asking participants to complete the CAPS questionnaire first and then using their scores as a pre-screening method to ensure a wider range of hallucination-free and hallucination-prone individuals are invited back to complete the main perceptual decision-making task. Alternatively, a comparison could be done between healthy controls and clinically diagnosed voice hearers.

Corlett and colleague's (2019) 'strong priors' theory of hallucinations posits that those who experience hallucinations hold inappropriate beliefs about the relative precision of incoming sensory signals and top-down beliefs. This theory hinges on the idea that agents form expectations at different levels of the cognitive hierarchy: not only expectations about the content of their experiences (i.e., *"I am likely to hear the word 'Pie' in this experiment"*), but also expectations about the precision of incoming sensory signals (i.e., *"I have heard a double beep, therefore I expect the next word to be clear"*). At first glance, our results may seem at odds with this theory; as participant's CAPS score increases their Expectation Effect decreases. However, what is important

to remember here is that 'Expectation Effect' in this instance refers to the extent to which participants are relying on expectations about the *sensory signal itself* (sensory precision prior) rather than forming expectations about *what* they are likely to experience (content prior).

Our results hint that hallucination prone individuals could be relying less on their expectations about sensory precision. A failure to accurately estimate the precision in low-level sensory systems could mean that perceivers end up relying more on top-down perceptual priors than they ought to. In this respect, our findings complement the 'strong prior' theory of hallucinations – since those prone to hallucinations could allocate an inappropriate weight to 'content expectations' (e.g., *"I expect to hear a voice"*) because they failed to estimate precision appropriately (e.g., *"I can't trust what I'm hearing, so I should rely on my expectations*").

Overall, these two experiments provide further support for influential Bayesian models of cognition, showing that listeners combine incoming evidence with prior expectations to estimate the precision of their hearing. However, more work is warranted in investigating the relationship between expectations about precision and unusual experiences such as hallucinations. Although we found a significant negative correlation between the two factors in one study, the same pattern of data did not reach significance in the second. Future work could explore this relationship further by looking at a more diverse groups of hallucinators (e.g., including clinically diagnosed voice hearers).

Chapter 5: Expected precision and evidence accumulation

5.1. Introduction

So far, work in this thesis has established that our subjective awareness is powerfully shaped by expectations about precision. However, an important question is whether the changes in perceptual awareness identified in previous chapters have consequences for metacognitive control – that is, adaptive behaviours which are thought to improve cognition and performance (Boldt et al., 2019; Boldt & Gilbert, 2022). Classical models of metacognition imagine an interconnected loop, where metacognitive monitoring mechanisms create subjective representations of uncertainty, and these representations are then used by metacognitive control mechanisms to coordinate overt behaviour (Nelson & Narens, 1990). A paradigmatic example of such monitoring is decision confidence. Metacognitive monitoring mechanisms can create feelings of confidence at the meta-level by tracking information in lower-level systems, and these meta-representations can then guide adaptive metacognitive control behaviours (Boldt et al., 2019; Boldt & Gilbert, 2022). Such behaviours include slowing down decisions (Yeung & Summerfield, 2012), manipulating our environment (Risko & Gilbert, 2016), seeking information (Desender et al., 2018) or asking for advice when we are uncertain (Bahrami et al., 2010; Shea et al., 2014). Our results so far suggest that expectations about precision can alter some kinds of metacognitive monitoring (e.g., subjective confidence) but is unclear whether these changes in confidence will also translate into changes in metacognitive control behaviour.

In this chapter we explore whether expectations about precision can influence the metacognitive control behaviour of evidence sampling (i.e., how long an individual takes to evaluate a situation before committing to a decision). Using a similar visual decision-making paradigm to that used in Chapters 2 and 3, we allowed participants to control their own sampling of the perceptual stimulus and investigated whether participants choose to slow down and gather more information when they expect the sensory world to be a more ambiguous place.

5.2. Experiment 9

5.2.1. Methods

Participants

Forty-two participants (30 female, 12 male, mean age = 35.4, SD = 8.61) were recruited via Prolific. All participants reported normal or corrected vision with no history of psychiatric or neurological illness. The sample size was selected to provide at least 80% statistical power to detect a medium-sized effect (Cohen dz = 0.453). This experiment was approved by the Research Ethics Committee at Birkbeck, University of London.

Participants were considered outliers if their individual effects (i.e., condition-wise differences in RTs) were >2.5SDs away from the sample mean. Similar to experiments in Chapter 2, we identified outliers based on participant's condition-wise effects (rather than specific condition means or trial-level data). Outliers were winsorized to values 2.5SDs away from the mean for inferential statistics, rather than adjusting raw datapoints. For Experiment 9, four participant's condition-wise effect for RTs were winsorized. Data points and their statistical significance were unchanged by this adjustment.

Procedure

Participants completed an online perceptual decision-making task programmed in PsychoPy (Peirce et al., 2019) adapted from Experiment 2. Participants were presented with clouds of right- or leftward moving dots and were asked to identify the overall direction of movement (i.e., left or right, see Figure 5.1). Each trial began with a fixation cross (500ms) before the presentation of a dot motion stimulus (max 5s). During stimulus presentation participants gave their response via keypress as to whether they thought the overall motion was to the left or right and their sampling time was recorded. In each motion cloud, a proportion of dots was programmed to move coherently left or coherently right, while the remaining dots moved in random directions.

Probabilistic cues (colours of fixation cross and stimulus dots) were the same as those used in Experiment 2 – allowing participants to form expectations about the strength of motion signals on

each trial and allowing us to investigate how such expectations bias perceptual judgements. Colour mappings were again counterbalanced across participants and participants were not explicitly informed about the association between the probabilistic cues and signal strength.

Like Experiment 2, the current experiment consisted of 560 trials. The first 160 acted as an initial training phase with perfectly deterministic mappings between colour and coherence to establish expectations about the cues (e.g., that 'blue cues' predicted clearer motion clouds). This was then followed by a 400-trial test phase where, like in Experiment 2, half of the trials *medium probe* trials. On these trials, participants received the same colour cues they had associated with clear or weak motion clouds but received an objective perceptual signal of intermediate strength (16% motion coherence). Given objective signal strength is identical on these trials, any differences in sampling times on these trials must reflect effects of expectations about signal strength – and these trials are the main trials of interest in analyses below.

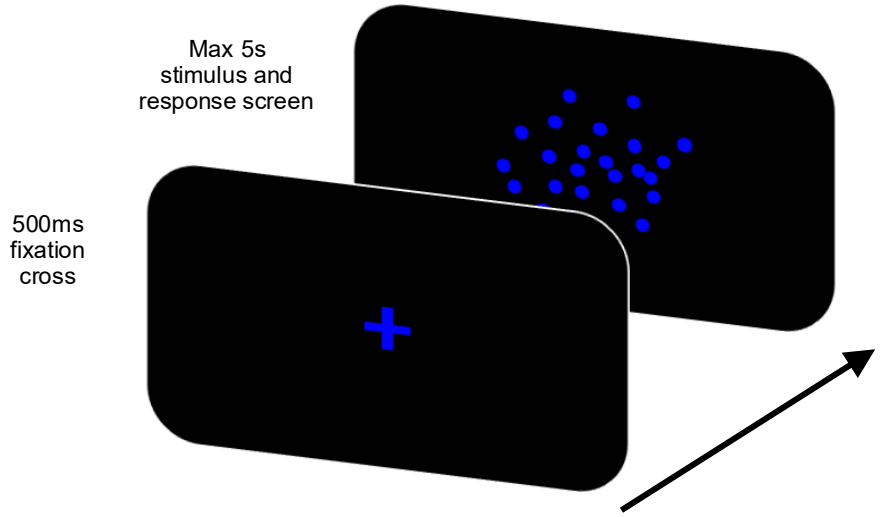


Figure 5.1 – Experimental task: Participants completed a motion perception task, deciphering the direction of left- or rightward moving dot cloud stimuli. Stimuli were displayed for a maximum of 5 seconds. Once a decision had been given via keypress the experiment would automatically move onto the next trial and sampling times were recorded. Colour cues manipulated expectations about the strength of motion patterns of each trial, e.g., if stimuli were blue participants could expect high motion coherence.

5.2.2. Results

We investigated how actual and expected precision influenced participant's metacognitive control behaviour; specifically, how long participants took to view the stimulus before committing to a perceptual decision.

As expected, participants accumulated more evidence from objectively weaker motion clouds than objectively stronger motion clouds – with longer RTs on truly weak (mean = 0.817ms, SD = 0.303) compared to truly strong trials (mean = 0.551ms, SD = 0.088, $t(41) = -7.272, p<0.001, dz = 0.197$). Critically though, we evaluated whether expectations about precision also affected evidence sampling by analysing the difference between RTs on *expect weak* and *expect strong*, specifically on the medium probe trials. The difference in RTs between the two trial types did not reach significance ($t(41) = -1.422, p=0.163, dz = 0.156,$ see Figure 5.2), however, we did see a consistent pattern across participants indicating that participants were slightly quicker to decipher the motion direction on ***expect strong*** trials (mean = 0.683s, SD = 0.187), than on *expect weak* trials (mean = 0.696s, SD = 0.189).

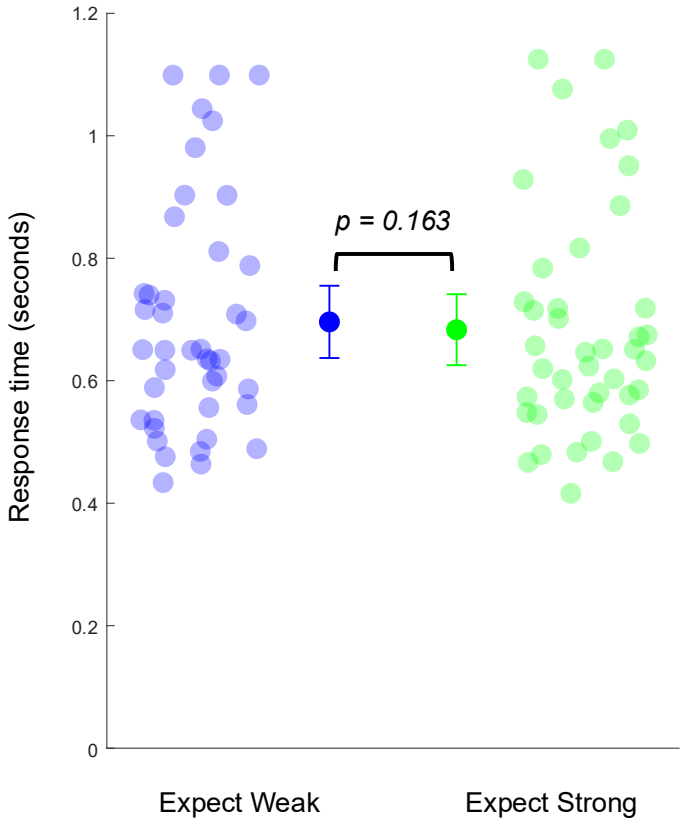


Figure 5.2 – Expected precision does not significantly alter participant sampling time: Participants took slightly longer to decipher left- and rightward movement of stimuli on *expect weak* trials, however, this difference did not reach significance. Errors bars represent 95% confidence intervals on the mean.

5.2.3. Discussion

In this chapter, we investigated whether expectations about precision could influence the metacognitive control behaviour of evidence sampling (i.e., how long individuals take to evaluate a stimulus before committing to a decision). It is already well documented that metacognitive or confidence judgements are closely linked with metacognitive control behaviours. For example, the more confident an individual feels about a decision the less likely they are to seek further information (Desender et al., 2019), ask for advice (Bahrami et al., 2010; Shea et al., 2014) or delay before committing to a decision (Yeung & Summerfield, 2012). In other chapters, we have found evidence that confidence and subjective awareness are changed by expectations about precision, but it remained unclear whether these changes would also translate into changes in

metacognitive control behaviours (i.e., *"I expect this signal to be ambiguous or weak, therefore I will adjust and take more time to sample the information before committing to a decision"*).

In Experiment 9, we tested this idea by allowing participants to control how long they sampled the perceptual stimulus before deciding on the overall motion direction of the stimulus. Critically, probabilistic cues allowed participants to form expectations about the strength or precision of the upcoming motion stimuli. Participant's sampling time would indicate whether they chose to slow down and gather more information when they *expected* the upcoming stimuli to be weak or imprecise. Our results showed that participants took marginally longer to report their decision when they expected weaker motion signals, in comparison to when they expected stronger signals – however, this pattern of results did not reach significance.

There are at least two possible interpretations of this pattern. The first is that this result is a 'false negative', due to limited statistical power. As the pattern of results is in the predicted direction (i.e., participants took longer to sample sensory evidence when they expected sensory signals to be weak), and perhaps this experiment would benefit from a larger sample and/or employ more sensitive measurements of sampling time (e.g., in the lab rather than online) to improve statistical power.

However, another possibility is that this is a 'true negative' result, and there is indeed a disconnect between the effects of expected precision on metacognitive monitoring and metacognitive control. If there is a disconnect, this may mean that the connection between subjective confidence and control behaviours is not as strong as usually thought. In other words, it could be possible to have a manipulation (like expectations) which makes an observer *feel* more confident, but which does not influence how they control their behaviour. This is precisely the question we target in the next chapter – where we investigate whether it is possible to dissociate the role of subjective uncertainty (confidence) and objective uncertainty (sensory precision) in the programming of control behaviours, like information seeking.

Chapter 6: Sensory uncertainty and information seeking

6.1. Introduction

So far, this thesis has largely focused on how expectations about precision influence metacognitive monitoring of perception. Throughout we have found evidence that our expectations about precision influence our subjective confidence and the subjective clarity of perceptual information. However, in the last chapter (Chapter 5) we found that these manipulations of expectation did *not* influence evidence sampling behaviour – a feature of our behaviour often thought to be connected to metacognitive feelings like confidence. This raises an intriguing possibility: that some kinds of adaptive control behaviours are largely influenced by *objective uncertainty* estimates, rather than subjective feelings. This would explain why changes in subjective confidence or subjective awareness may not always be translated into changes in adaptive control. Here, we investigate this in the context of information seeking.

Humans ask questions in talks and read papers in journals to *seek information.* Other animals seek information too, like when primates look inside opaque containers before making decisions about which one they'd prefer. In humans and other creatures, gathering information allows us to improve our internal models of the outside world (e.g., *"which scientific theories are true?"* or *"which tube contains a peanut?"*) which in turn leads to adaptive cognition and choice.

In humans, information seeking is generally thought to depend on *explicit metacognition*: introspective processes that subjectively monitor the uncertainty in our own mental states, like subjective feelings of confidence (Fleming et al., 2012; Yeung & Summerfield, 2012). Elegant experiments have shown that factors which manipulate subjective confidence, without changing objective accuracy, can also alter decisions to seek further information. For example, sampling more variable sensory evidence causes us to feel less confident in our perceptual decisions (Boldt et al., 2017; Gardelle & Mamassian, 2015), and increases the odds that we opt for a 'clearer look' at a stimulus before committing to a choice (Desender et al., 2018, 2019). The logic here is that the

subjective experiences of uncertainty we feel attached to our decisions are instrumental in causing our actions to seek information. We seek more information *because* we don't feel confident.

The close connection between subjective metacognition and information seeking in adult humans has inspired researchers across the cognitive sciences to use information seeking as a window into the metacognitive abilities of creatures and populations that cannot explicitly comment on their own mental states. For instance, evidence that nonhuman animals seek information in more ambiguous environments has been taken as evidence that these creatures can consciously monitor their knowledge states in the same way that adult humans can (Call & Carpenter, 2001; Hampton et al., 2004). At the same time, evidence that preverbal infants can seek help in the face of more uncertain decisions has been taken as a sign that preverbal infants "consciously experience their own uncertainty" (Goupil et al., 2016).

This scientific strategy is intuitive, and there is good evidence that information seeking in infants and animals is genuinely sensitive to uncertainty. However, a neglected nuance in these areas is that our minds and brains can be sensitive to *uncertainty* in ways that do not entail *metacognition.* For example, Bayesian models of the brain suggest that uncertainty or 'precision' is a property that could be estimated and represented at every level of the brain's hierarchy – but not all of these uncertainty representations enter into subjective awareness or are used for metacognitive control. For instance, there is compelling evidence that low-level multisensory integration (e.g., combining vision and touch) depends on computations that estimate the uncertainty in sensory signals (Alais & Burr, 2004; Ernst & Banks, 2002), but we are typically only aware of the result of these uncertainty computations (i.e., the multisensory percept) rather than the uncertainty estimates themselves (Deroy et al., 2016).

The existence of such 'subpersonal' uncertainty estimates even in the adult human brain makes it possible that information seeking action could indeed be driven by uncertainty – but this may not involve the same uncertainty computations that generate metacognitive feelings like confidence (see Figure 6.1). A potential dissociation between subjective metacognition and information seeking would undermine the idea that information seeking necessarily reflects conscious introspection about a creature's own mind.

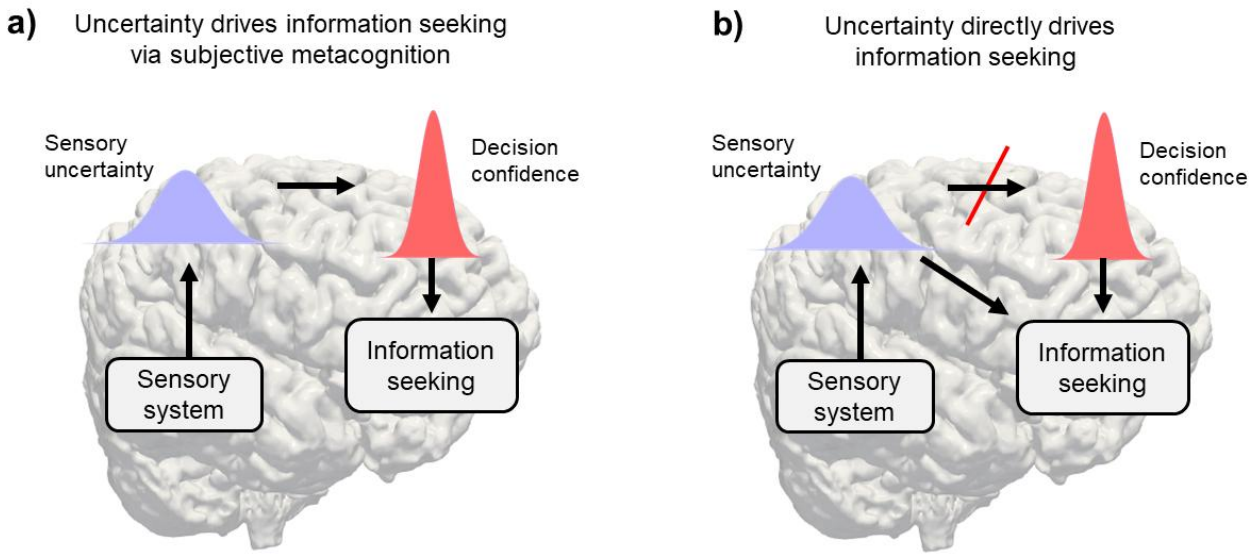


Figure 6.1: Uncertainty, metacognition and information search. a) According to metacognitive theories there is a tight connection between information seeking and subjective metacognitive states (e.g., feelings of confidence about our perceptual choices). In this way of thinking, internal estimates of sensory evidence strength (e.g., of the kind found in the parietal cortex - Bang & Fleming, 2018) are one source of evidence used to construct subjective feelings of confidence (e.g., in dorsolateral prefrontal cortex, Shekhar & Rahnev, 2018). These subjective feelings then guide control behaviours, which is why we search for more information when confidence is low (Desender et al., 2018). **(b)** However, an alternative possibility is that uncertainty estimates in the brain *directly* control information seeking without the involvement of introspective metacognitive computations e.g., we may seek information when sampled signals are less precise. If this second possibility is true, breaking the link between sensory uncertainty and decision confidence should reveal situations where sensory ambiguity drives information seeking, even if subjective confidence has not changed. (NB: Imagined neural locations are speculative)

Here we reveal just such a dissociation. Across two experiments, we have participants make perceptual choices while we independently manipulated two separate kinds of uncertainty: uncertainty caused by sampled perceptual evidence (sensory uncertainty) and uncertainty caused by choice boundaries (decisional uncertainty). Across these experiments we find that it is sensory (and not decisional) uncertainty which drives information seeking behaviour.

Since subjective confidence is usually thought to be closely connected to decision difficulty, but not sensory uncertainty (see Bang & Fleming, 2018), this pattern suggests that information seeking can operate independently of the mechanisms involved in subjective metacognitive monitoring. This may in turn mean that some aspects of our behaviour are not controlled by subjective estimates of precision, contrary to what is widely assumed.

6.2. Experiment 10

Participants completed a perceptual decision-making task judging the prominent direction of motion in clouds of moving dots in relation to a comparison line (i.e., did the cloud move clockwise or counterclockwise relative to the comparison). The reference line remained on screen until participants registered their decision or chose to have a second look at the stimuli. Points could be won for correct answers or deducted for incorrect answers. Sensory and decisional uncertainty were manipulated independently of each other to probe which kind of uncertainty influenced the probability of participants seeking more information (having a second look at the stimuli).

6.2.1. Methods

Participants

Fifteen participants (12 female, 3 male, mean age = 32.2, SD = 8.92) completed Experiment 10 – this sample size was chosen arbitrarily. All participants reported normal or corrected vision, with no history of psychiatric or neurological illness. Participants were recruited via local databases and tested in person at Birkbeck. The experiment was approved by the Research Ethics Committee at Birkbeck University of London.

Participants were considered outliers if their probability of seeking more information was >2.5SDs away from the mean. Outliers were winsorized to values 2.5SDs away from the mean for inferential statistics. Data patterns and their statistical significance were unchanged by this adjustment.

Procedure

Participants completed a perceptual decision-making task in PsychoPy (see Figure 6.2). Each trial began with a fixation cross (1500ms) followed by a cloud of moving dots (300ms) and a blank screen (700ms). These clouds were programmed to move coherently in a direction randomly chosen between 0 – 180 degrees (i.e., the top half of an imaginary circle). After viewing the moving dots, participants saw a comparison line and were required to judge whether the cloud had moved clockwise or counterclockwise relative to the comparison. The reference line remained on screen until participants registered their decision with a key press.

In the task we independently manipulated two kinds of uncertainty: *sensory uncertainty* and *decisional uncertainty*. We manipulated sensory uncertainty by creating two sensory evidence conditions – *strong* and *weak.* Motion coherence was lower on *weak* trials (16 %) than on *strong* trials (48 %). We independently manipulated decisional uncertainty by using adaptive staircases to titrate the decision line (Leek, 2001). For *easier* trials, staircases targeted a level where accuracy would be ~70%, using a 1-up-2-down adjustment rule – meaning that if participants made two correct decisions in a row the reference line was drawn closer to true motion direction (making choices harder) while if participants made one error the line was adjusted to be drawn further (making choices easier). For *harder* conditions staircases targeted an accuracy level ~50%, instead using a 1-up-1-down rule – meaning that the reference line was adjusted to be harder after each correct choice and easier after each error. Crucially, these staircases were run independently for each of the four experimental conditions (*Strong Easy, Strong Hard, Weak Easy, Weak Hard*), and these staircases were left running throughout the experiment. This made it possible to create conditions where sensory uncertainty was decoupled from decision difficulty (e.g., by setting different decision boundaries for *strong* and *weak* motion clouds that would yield accuracies of ~70% in *Strong Easy* and *Weak Easy* conditions).

In all experiments, participants completed this perceptual task in three blocks or ‘phases’. The first two phases were identical in each experiment. The first was a practice block (80 trials) where participants were familiarised with the task, but did not gain any points for the decisions they made. This phase also allowed the independent adaptive staircases to converge on *Easy* and *Hard*

difficulty levels for each of the four conditions. In the second block (200 trials) participants were informed that they would now earn points for making correct decisions, and that these points would be used to calculate a bonus payment at the end of the experiment. In this phase, participants received +4 points for a correct choice and +0 points for an error. This block primarily served to allow us to verify that our task manipulation successfully dissociated main effects of sensory uncertainty and decisional uncertainty on objective task performance.

However, the third and final block (200 trials) was the primary block of interest; here participants had the option to seek more information before making a choice. If participants opted to 'look again' at the stimulus, they saw the same motion stimulus (at 100% coherence) and then made the same discrimination decision – but for a reduced reward. Participants received +5 points for a correct choice made without an extra look, +2 points for a correct choice made after looking again or +0 points for an incorrect decision. Breaks were offered every 20 trials.

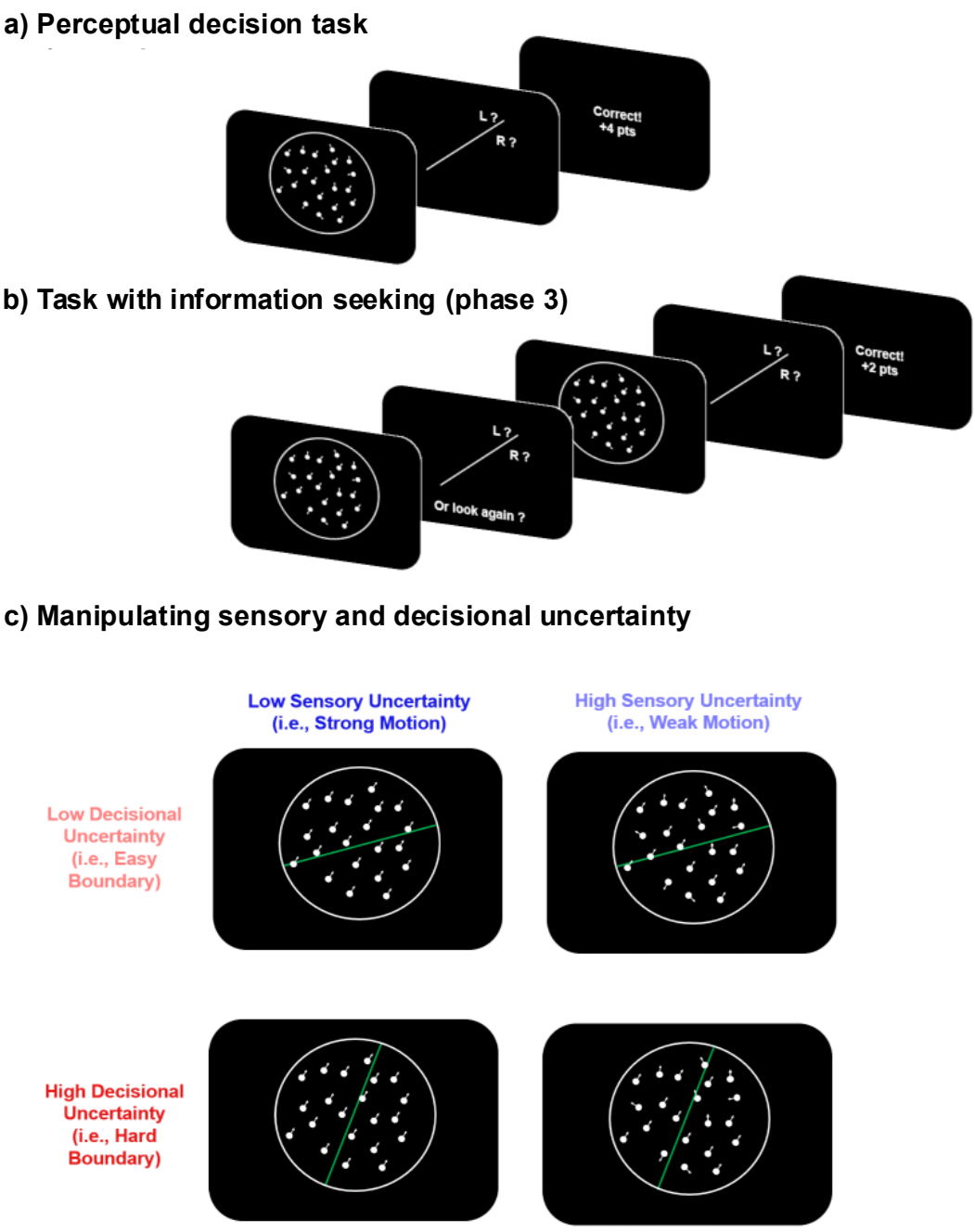


Figure 6.2: Experimental task and manipulations. a) Participants judged the predominant direction of motion in moving dot clouds and earned points for correct decisions. **b)** In the third phase of the experiment participants could seek more information (another 'look') before committing to a decision **c)** We independently manipulated uncertainty in the stimulus and uncertainty in the choice to evaluate how these distinct kinds of uncertainty influence information seeking and subjective confidence.

6.2.2. Results

Dissociating sensory and decision uncertainty

Before analysing our main variable of interest (information seeking) we first evaluated whether our experimental design was successful in dissociating sensory uncertainty from decisional uncertainty. Participants first completed a practice phase to begin the adaptive staircasing, followed by a main block of 200 trials where participants made these perceptual decisions across four conditions (*weak easy, weak hard, strong easy, strong hard).* A 2x2 within-subjects ANOVA with sensory uncertainty (*strong* or *weak* visual evidence) and decisional uncertainty (*easy* or *hard* choice boundaries) was used to investigate this question.

Analysing decisions in this phase of the experiment revealed that our design was successful in decoupling sensory and decisional uncertainty. Average accuracy was substantially higher on *easier* trials (mean = 0.720, SD = 0.081) than on *harder* trials (mean = 0.575, SD = 0.063) – $F(1, 14) = 57.396$, $p<0.001$, $\eta_p^2 = 0.804$ – suggesting that our staircasing procedure was successful.

Despite staircasing, there remained a statistically significant difference in choice accuracy between trials with *weak* (mean = 0.638, SD = 0.07) compared to *strong* visual evidence (mean = 0.657, SD = 0.074; $F(1, 14) = 4.759$, $p = 0.047$, $\eta_p^2 = 0.254$) – however, this difference was very small. Critically though, these two factors did not interact ($F(1, 14) = 1.256$, $p = 0.281$, $\eta_p^2 = 0.082$) – meaning that our experimental paradigm successfully dissociated two sources of uncertainty that could affect information seeking and metacognition. A full breakdown of mean and SDs for accuracies across trial types are given in table 6.1.

	Weak-Easy	Weak-Hard	Strong-Easy	Strong-Hard
Mean	0.703	0.572	0.736	0.577
SD	0.08	0.059	0.081	0.067

Table 6.1: Full breakdown of task accuracy means and SDs across trial types. Trials could either be *weak* or *strong* in terms of 'sensory uncertainty' and *easy* or *hard* in terms of 'decisional uncertainty', creating four unique trial types. Here we present the mean accuracy and accompanying SD for each of these four trial types.

Sensory (not decisional) uncertainty controls information seeking

Participants proceeded to a final block of 200 trials where they had the option to *seek information*. Before committing to a choice, observers could push a button to obtain a 'clearer look' at the stimulus (an additional stimulus presentation, where the same motion direction was shown at 100% coherence).

However, the more intriguing question here is what sort of uncertainty (decisional or sensory) has more of an impact on information seeking behaviour. If information seeking depends on an introspective appraisal of decision accuracy (e.g., "probability correct" – (Guggenmos, 2022; Pouget et al., 2016) we may expect observers to seek more information on trials where they are more likely to make errors. However, our analyses revealed no main effect of decision difficulty (*easier* vs *harder*) on information seeking decisions ($F(1, 14) = 2.475$, $p = 0.138$, $\eta_p^2 = 0.15$ – see Figure 6.3).

In contrast, information seeking was strongly shaped by sensory uncertainty ($F(1,14) = 9.179$, $p = 0.009$, $\eta_p^2 = 0.396$), with a higher probability of seeking more information on *weak* trials (mean = 0.142, SD = 0.10) than on *strong* trials (mean = 0.087, SD = 0.072). This means that observers sought more information when sensory evidence was more ambiguous – even though there were many ambiguous stimuli where the probability of a correct decision was relatively high, and many clear stimuli where the probability of a correct decision was relatively low. Experiment 10 hinted at

a possible interaction between sensory and decisional uncertainty, but this did not reach significance ($F(1,14) = 4.441$, $p=.054$, $\eta_p^2 = 0.241$).

In sum, these results show that – when sensory uncertainty and decisional uncertainty are decoupled – it is the ambiguity in the sampled sensory signals rather than the difficulty of the choice that controls information seeking action.

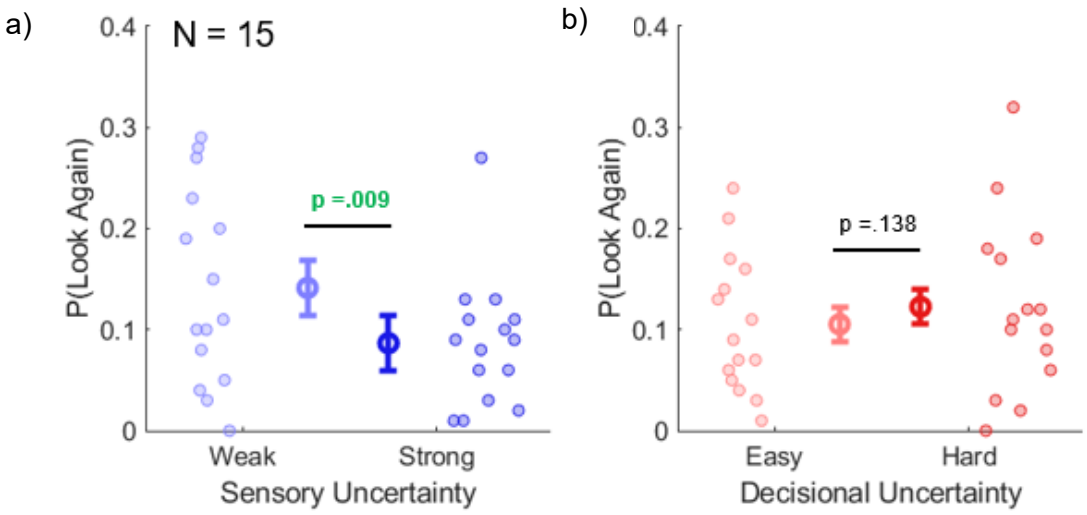


Figure 6.3: Sensory (but not decisional) uncertainty controls information seeking: a) We found that information seeking behaviour (decisions to 'look again') were strongly shaped by the ambiguity in sensory signals, **b)** but not influenced by manipulations of decision difficulty (i.e., easier or harder decision boundaries). Hollow markers indicate sample means in each condition, and error bars denote 95% within-subjects confidence intervals on the difference between conditions.

6.2.3. Discussion

Here we find that information seeking behaviour is heavily influenced by sensory uncertainty, but *not* by subjective decisional uncertainty. When sensory uncertainty was high (weak or imprecise sensory signals), participants were significantly more likely to opt for a second look at the stimuli, in comparison to when sensory uncertainty was low (strong or precise sensory signals). However, probability of having a second look at the stimulus was not significantly affected by subjective decisional uncertainty. This result may accord well with results in Chapter 5, where a factor which

affected subjective uncertainty (i.e. expected precision) did not significantly influence the control of perceptual evidence accumulation (which could be construed as a form of 'information seeking' in its own right).

6.3. Experiment 11

Following Experiment 10, a replication was conducted with an increased sample size to see whether our initial findings were robust.

6.3.1. Methods

Participants

A new sample of 38 participants (24 female, 14 male, mean age= 32.2, SD = 8.10) were recruited via Prolific. This sample size was chosen to provide at least 80% power to detect the possible effect of decision difficulty that did not reach significance in Experiment 10 (η_p^2 = 0.15). The same exclusion and outlier identification procedures were used as in Experiment 10. Data patterns and their statistical significance were unchanged by this adjustment.

Procedure

The stimuli, paradigm and procedure were exactly the same as that used in Experiment 10.

6.3.2. Results

We carried out exactly the same analysis as Experiment 10.

Dissociating sensory and decisional uncertainty

We first evaluated whether our experimental design was again successful in dissociated sensory uncertainty from decisional uncertainty. Again, average accuracy was substantially higher on *easier*

trials (mean = 0.683, SD = 0.115) than on harder trials (mean = 0.572, SD = 0.075; *F(1, 37) = 69.231, p<0.001, η_p^2 = 0.425)*, suggesting that our staircasing procedure was successful. We also, again, found a small but consistent difference in choice accuracy between trials between *weak* (mean = 0.607, SD = 0.092) compared to *strong* visual evidence (mean = 0.648, SD = 0.098; *F(1, 37) = 12.472, p = 0.001).* These factors did not interact (*F(1, 37) = 0.50, p = 0.484, η_p^2 = 0.002),* meaning that our experimental paradigm successfully dissociated the two sources of uncertainty. A full breakdown of mean and SDs for accuracies across trial types are given in table 6.2.

	Weak-Easy	**Weak-Hard**	**Strong-Easy**	**Strong-Hard**
Mean	0.665	0.548	0.700	0.596
SD	0.117	0.066	0.113	0.083

Table 6.2: Full breakdown of task accuracy means and SDs across trial types. Trials could either be *weak* or *strong* in terms of 'sensory uncertainty' and *easy* or *hard* in terms of 'decisional uncertainty', creating four unique trial types. Here we present the mean accuracy and accompanying SD for each of these four trial types.

Sensory (not decisional) uncertainty controls information seeking

Consistent with Experiment 10, we found no main effect of decision difficulty (*easier* vs *harder*) on information seeking decisions (*F(1, 37) = 2.053 , p = 0.160 , η_p^2 = 0.005* – see Figure 6.4). Yet again, we found that information seeking was strongly influenced by sensory uncertainty *(F(1, 37) = 14.546, p<0.001, η_p^2 = 0.233)*, with a higher probability of seeking more information on *weak* trials (mean = 0.087, SD = 0.104) than on *strong* trials (mean = 0.038, SD = 0.055). There was no interaction between sensory and decisional uncertainty (*F(1,37) = 0.910, p = 0.346, η_p^2 = 0.002)*.

These results replicate Experiment 10; suggesting that it is sensory ambiguity rather than choice difficulty that controls information seeking action.

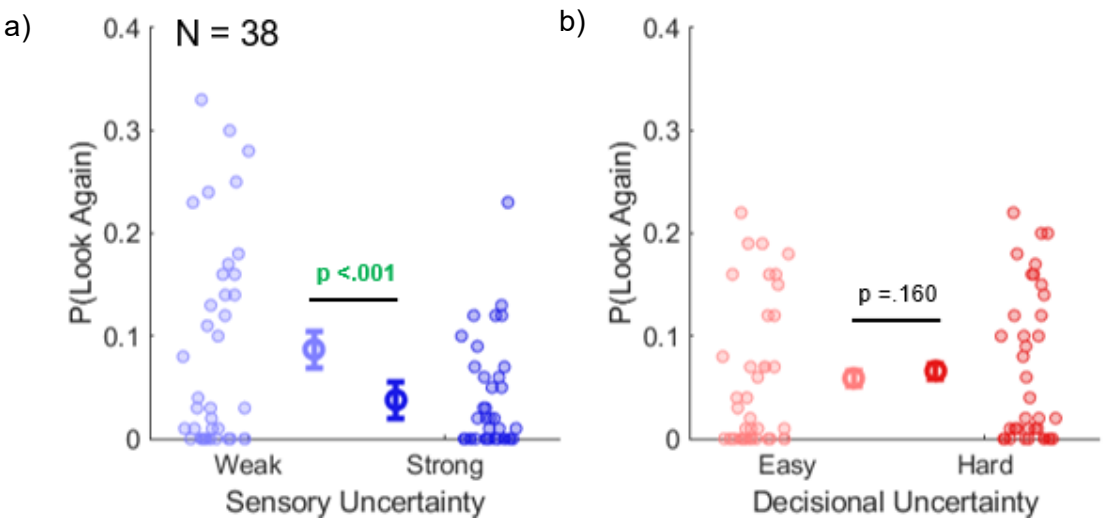


Figure 6.4: Sensory (but not decisional) uncertainty controls information seeking: a) We found that information seeking behaviour (decisions to 'look again') were strongly shaped by the ambiguity in sensory signals, **b)** but not influenced by manipulations of decision difficulty (i.e., easier or harder decision boundaries). Hollow markers indicate sample means in each condition, and error bars denote 95% within-subjects confidence intervals on the difference between conditions.

6.3.3. Discussion

Results from Experiment 11 replicated those of Experiment 10; information seeking behaviour was significantly influenced by sensory uncertainty but *not* decisional uncertainty. When sensory uncertainty was high (weak or imprecise sensory information), participants were much more likely to opt for a second look at the stimuli, in comparison to when sensory uncertainty was low. These results go against the traditional idea that information seeking depends on conscious metacognitive computations (i.e., confidence judgements).

6.4. General Discussion

Metacognitive theories suggest there is a tight connection between information seeking and subjective confidence. In these theories, decisions to seek more information are driven by the subjective feelings of uncertainty attached to our choices. Indeed, the connection between

information seeking and explicit feelings like subjective confidence is assumed to be so strong that information seeking is often used by researchers in developmental psychology or comparative cognition as a window into the conscious introspective states of preverbal children or nonverbal animals. However, here we find that a manipulation usually thought to influence subjective uncertainty (decision difficulty) did not affect information seeking, while a more objective form of sensory ambiguity did (sensory uncertainty).

This dissociation undermines the idea that information seeking necessarily depends on conscious metacognition computations – of the kinds that create subjective feelings of confidence. If this were true, factors influencing subjective confidence about our perceptual choices should also shape information seeking (Desender et al., 2018), but we do not see this here.

Instead, our results are consistent with an emerging picture from computational neuroscience, where a variety of uncertainty estimates are stored at different levels throughout the hierarchy of the mind and brain, but not all forms of uncertainty are consciously accessible (Pouget et al., 2016; Yon & Frith, 2021). For instance, in Bayesian models, brains keep track of the uncertainty or 'precision' in sensory circuits to control the integration of information across the senses (Ernst & Banks, 2002) or the combination of momentary evidence with existing prior beliefs (Yon, 2021). Our results are consistent with the idea that similar representations of sensory uncertainty might be used to guide information seeking action – such that humans and other animals might explore more in more ambiguous sensory environments – but without the involvement of explicit metacognition.

Our results suggest that some kinds of uncertainty can influence information seeking without mediation through changes in subjective confidence, possibly utilising a pathway as outlined in Figure 6.1. However, this does not imply that metacognition and information seeking *never* interact. For example, it remains highly plausible that metacognitive introspection can lead us to seek more information before committing to a perceptual choice. Indeed, Desender and colleagues (2018) found that information seeking was driven more so by evidence reliability, not because the factors were directly connected, but because of evidence reliability's inextricable link to confidence, which both feed into information seeking behaviour. More broadly there is also good evidence connecting

changes in subjective confidence to other forms of metacognitive control – such as changes in evidence accumulation (Balsdon et al., 2020), cognitive offloading (Boldt & Gilbert, 2019; Scott & Gilbert, 2024) and even curiosity about metacognition itself (Recht et al., 2024). However, our results do call into question whether information seeking and subjective are as tightly connected as traditionally thought. Our results reveal a possible alternative: that uncertainty may be able to alter information seeking behaviour *without* altering explicit metacognition.

Chapter 7: Discussion

This thesis has concerned influential Bayesian models of the mind. Bayesian models of perception suggest that observers estimate the precision of incoming evidence and use these estimates to decide how to combine information from different sensory systems (Ernst & Banks, 2002) or how to combine incoming evidence and prior expectations (Yon & Frith, 2021) – giving more weight to incoming signals that are currently most precise. At the same time, internal estimates of precision are thought to be fundamental in how the brain metacognitively monitors its own uncertainty – creating subjective feelings like confidence (Geurts et al., 2022).

An important shift in contemporary Bayesian models is the idea that precision is not estimated on incoming evidence alone but also assumes that agents form probabilistic beliefs about how precise information is *likely* to be, and these expectations are incorporated into precision estimates (Friston, 2018).

This idea of *expected precision* has become increasingly embedded in theoretical models of the Bayesian brain and often used to explain unusual experiences such as hallucinations (Corlett et al., 2019). Forming beliefs about precision would help agents to estimate uncertainty – which may often be difficult to compute (Yon & Frith, 2021) – and also optimise metacognitive monitoring of perception and action. However, while we can potentially explain various aspects of perception and metacognition by assuming agents form beliefs about precision, it was previously unclear how or whether expectations about perceptual precision are actually formed and the role these expectations play in cognition and behaviour.

This thesis has provided important insights into these questions. In this Discussion section, I summarise the key findings and outline questions for future research, spread across three themes: *fundamental mechanisms of expected precision, expected precision and atypical experiences* and *metacognitive monitoring and metacognitive control*

7.1. Fundamental mechanisms of expected precision

In **Chapter 2,** we provided the first empirical support for the idea that probabilistic expectations about precision influence subjective confidence and subjective awareness. Moreover, we present a new predictive learning model which can explain how these effects of expectations on awareness can arise. We extend this work in **Chapter 4** to reveal comparable influences in audition as well as vision. Taken together, our results provide support for contemporary Bayesian models of the mind, and the idea that agents do not only 'read out' the reliability of information arriving at their senses but also take into account prior knowledge about how reliable or 'precise' information is likely to be.

We extend this work into neural mechanisms in **Chapter 3**. Here we found representations of sensory uncertainty in the insula and area MT, but only find these representations are modulated by expectations in the insula. These findings point to the possibility that the insula plays a role in encoding 'precision prediction errors' that capture the difference between the precision we are expecting and the precision we receive. These prediction errors may be crucial in the formation of expectations about precision we see in Chapter 2.

Unravelling these neural and computational mechanisms is an important area for future work. One possibility could be to test more explicitly the 'prediction error' hypothesis about the role of the insula in precision learning. This could be achieved by using the predictive learning model introduced in Chapter 2 to estimate trial-by-trial prediction errors, testing whether the error signals implied by the model are indeed encoded in the same insula region. It would also be valuable to intervene on activity in this insula region to see whether this affects the acquisition and use of precision expectations. This could involve the use of transcranial magnetic stimulation (TMS) to disrupt neural activity in this area (Bolognini & Ro, 2010), or more profitably transcranial ultrasound (TUS) which are more effective at targeting deeper cortical structures (Kubanek, 2018). Presumably, if the insula is involved in the tracking of sensory uncertainty *and* the formation of expectations about precision, disrupting the neural function here would disrupt both processes. This could possibly lead to lower accuracy on the task (i.e. more random ratings of subjective clarity) but could also prevent the acquisition of expectations about precision – abolishing the

effects we see in our behavioural paradigms.

7.2. Expected precision and atypical experiences

The concept of 'expected precision' is central to Bayesian models of atypical experiences – like hallucinations. In **Chapter 4** we investigated whether the expected precision mechanisms identified in previous chapters may be connected to unusual experiences such as hallucinations. We found some evidence that those prone to hallucinations were less able to form or use expectations about sensory precision when estimating the reliability of the sensory world. A failure to accurately estimate the precision in low-level sensory systems could mean that perceivers end up relying on top-down perceptual priors more than they ought to, which could engender hallucination-like experiences. These findings therefore provide general support for the application of Bayesian ideas and 'expected precision' to atypical experiences – like those which characterise psychotic illness.

However, associations between precision expectations and hallucinations were not consistent across our studies (i.e., significant in one case, and non-significant in another). There are at least two possible ways this line of work could be extended in future. One possibility could be to recruit a more diverse sample from the general population, actively stratifying the range of unusual experience scores (e.g., the CAPS questionnaire) to improve the sensitivity of the correlational analyses, capturing a wider spectrum of hallucination proneness. A second complementary approach would be to run the same study with a case-control design, comparing healthy controls and a clinical sample of voice-hearers. This could potentially reveal the extent to which expected precision influences perceptual awareness and how extensively the neural mechanisms we have identified in previous chapters are disrupted in psychotic illness.

7.3. Metacognitive monitoring and metacognitive control

In the final two chapters, we focused on whether expected precision and sensory uncertainty have any consequences for metacognitive control. In **Chapter 5** we probed evidence accumulation

(sampling time). We found that participants took slightly longer to make a decision when they *expected weak* signals in comparison to when they *expected strong* signals – however this pattern of results did not reach significance. This result could possibly be a false negative, as the pattern of results is in the predicted direction and maybe with a larger sample and more sensitive sampling time measurement, this may reveal a clearer result. Alternatively, these results could be a true negative, and there is a disconnect between the effects of expected precision on metacognitive monitoring and metacognitive control. This may mean that the connection between subjective confidence and control behaviours is not as strong as initially thought.

To draw a more solid conclusion about the involvement of expected precision in metacognitive control, future research should address the methodological suggestions outlined above (i.e., a larger sample size and lab-based testing to allow for more sensitive time measurement). After repeating the original analyses of this proposed data, drift diffusion modelling could then be introduced as an additional analysis. Drift diffusion modelling follows an agent's decision-making process as they accumulate information over time until the process hits a certain threshold and a decision is ultimately made (Fudenberg et al., 2020). Such modelling techniques can yield measurements such as the drift rate and decision thresholds. Drift rate reflects the speed of evidence accumulation towards the decision threshold (Fudenberg et al., 2020). Decision thresholds indicate how much evidence is needed to make a decision – higher thresholds lead to longer sampling times and more cautious decisions (Myers et al., 2022). Identifying and comparing these parameters between 'expect weak' and 'expect strong' conditions in a replication of our evidence accumulation experiment would allow a more granular investigation of the processes involved in metacognitive control of continuous decisions than looking at response times alone. For example, decision thresholds and/or drift rates may vary between the different precision expectation conditions, but these possibilities cannot be disentangled looking at sampling time alone.

We also investigated a second form of metacognitive control: information seeking. Our results from Chapter 5 raises an intriguing possibility: that some kinds of adaptive control behaviours are largely influenced by *objective uncertainty* estimates, rather than subjective feelings (confidence or

expectations). Therefore, the experiments in Chapter 6 investigated whether sensory uncertainty alone would influence participant's probability of seeking information, rather than just changes in subjective confidence. Here we find that sensory uncertainty, but not decisional uncertainty significantly increases people's probability of seeking more information to make a decision. This is surprising, given that decisional uncertainty is usually closely connected to subjective confidence and suggested that decisions to seek information and metacognitive feelings do not always rely on the same cognitive computations.

One way that this possibility could be tested more directly in future work could be to combine metacognitive monitoring and control into a single experiment. For instance, future studies could extend the 'expected precision' conditioning task used throughout this thesis to collect both information seeking ('second look') and subjective clarity ratings simultaneously. Such an experimental structure would allow us to test more directly whether expectations are being formed to shape subjective awareness separately from shaping overt behaviour, or if awareness and behaviour are influenced similarly when both are probed together.

7.4. Conclusion

In sum, this thesis provides support for Bayesian models of expected precision. We provide behavioural evidence to support these ideas and reveal possible neural mechanisms underlying the process of learning and predicting precision. We also provide tentative support for theories that implicate aberrant expectations about precision in the genesis of unusual experiences like hallucinations. Though our results leave open some questions about the relationship between expectations, metacognitive monitoring and metacognitive control, this thesis provides important empirical groundwork for Bayesian theories of the mind and also suggests intriguing possibilities for future directions in this area of research.

References

Aitchison, L., Bang, D., Bahrami, B., & Latham, P. E. (2015). Doubly Bayesian Analysis of Confidence in Perceptual Decision-Making. *PLOS Computational Biology*, *11*(10), e1004519. https://doi.org/10.1371/journal.pcbi.1004519

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. (2010). Optimally interacting minds. *Science*, *329*(5995), 1081–1085. https://doi.org/10.1126/science.1185718

Balsdon, T., Wyart, V., & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-15561-w

Bang, D., Ershadmanesh, S., Nili, H., & Fleming, S. M. (2020). Private–public mappings in human prefrontal cortex. *eLife*, *9*(e56477).

Bang, D., & Fleming, S. M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(23), 6082–6087. https://doi.org/10.1073/pnas.1800795115

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

Bell, V., Halligan, P. W., & Ellis, H. D. (2006). The Cardiff Anomalous Perceptions Scale (CAPS): A New Validated Measure of Anomalous Perceptual Experience. *Schizophrenia Bulletin*, *32*(2), 366–377. https://doi.org/10.1093/schbul/sbj014

Boldt, A., de Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *Journal of Experimental Psychology. Human Perception and Performance*, *43*(8), 1520–1531. https://doi.org/10.1037/xhp0000404

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, *4*(1), 45. https://doi.org/10.1186/s41235-019-0195-y

Boldt, A., & Gilbert, S. J. (2022). Partially overlapping neural correlates of metacognitive monitoring and metacognitive control. *Journal of Neuroscience*, *42*(17), 3622–3635. https://doi.org/10.1523/jneurosci.1326-21.2022

Boldt, A., Schiffer, A. M., Waszak, F., & Yeung, N. (2019). Confidence predictions affect performance confidence and neural preparation in perceptual decision making. *Scientific Reports*, *9*(1), 1–17. https://doi.org/10.1038/s41598-019-40681-9

Bolognini, N., & Ro, T. (2010). Transcranial Magnetic Stimulation: Disrupting Neural Activity to Alter and Assess Brain Function. *The Journal of Neuroscience*, *30*(29), 9647–9650. https://doi.org/10.1523/JNEUROSCI.1990-10.2010

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, *28*, 157–189. https://doi.org/10.1146/annurev.neuro.26.041002.131052

Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition*, *3*(4), 207–220. https://doi.org/10.1007/s100710100078

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., & Powers, A. R. (2019). Hallucinations and strong priors. *Trends in Cognitive Sciences*, *23*(2), 114–127. https://doi.org/10.1016/j.tics.2018.12.001

Daniel, R., & Pollmann, S. (2012). Striatal activations signal prediction errors on confidence in the absence of external feedback. *NeuroImage*, *59*(4), 3457–3467. https://doi.org/10.1016/j.neuroimage.2011.11.058

Davidson, M. J., Macdonald, J. S. P., & Yeung, N. (2022). Alpha oscillations and stimulus-evoked activity dissociate metacognitive reports of attention, visibility, and confidence in a rapid visual detection task. *Journal of Vision*, *22*(10), 20. https://doi.org/10.1167/jov.22.10.20

Davies, D. J., Teufel, C., & 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. https://doi.org/10.1093/schbul/sbx177

Dayan, P., & Kakade, S. (2000). Explaining Away in Weight Space. *Advances in Neural Information Processing Systems*, *13*. https://proceedings.neurips.cc/paper/2000/hash/eb1e78328c46506b46a4ac4a1e378b91-Abstract.html

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

Deroy, O., Spence, C., & Noppeney, U. (2016). Metacognition in Multisensory Perception. *Trends in Cognitive Sciences*, *20*(10), 736–747. https://doi.org/10.1016/j.tics.2016.08.006

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761–778. https://doi.org/10.1177/0956797617744771

Desender, K., Murphy, P., Boldt, A., Verguts, T., & Yeung, N. (2019). A postdecisional neural marker of confidence predicts information-seeking in decision-making. *Journal of Neuroscience*, *39*(17), 3309–3319. https://doi.org/10.1523/JNEUROSCI.2620-18.2019

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

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

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, *15*(1), 146–154. https://doi.org/10.1038/nn.2983

Fleming, S. M. (2017). HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, *2017*(1), nix007. https://doi.org/10.1093/nc/nix007

Fleming, S. M. (2020). Awareness as inference in a higher-order state space. *Neuroscience of Consciousness*, *2020*(1), niz020. https://doi.org/10.1093/nc/niz020

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91–114. https://doi.org/10.1037/rev0000045

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1338–1349. https://doi.org/10.1098/rstb.2011.0417

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: Computation, biology and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1280–1286. https://doi.org/10.1098/rstb.2012.0021

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2016). Metacognition about the past and future: Quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness*, *2016*(1), niw018. https://doi.org/10.1093/nc/niw018

Friston, K. (2018). Does predictive coding have a future? *Nature Neuroscience*, *21*(8), 1019–1021. https://doi.org/10.1038/s41593-018-0200-7

Fudenberg, D., Newey, W., Strack, P., & Strzalecki, T. (2020). Testing the drift-diffusion model. *Proceedings of the National Academy of Sciences*, *117*(52), 33141–33148. https://doi.org/10.1073/pnas.2011446117

Gardelle, V. de, & Mamassian, P. (2015). Weighting Mean and Variability during Confidence Judgments. *PLOS ONE*, *10*(3), e0120870. https://doi.org/10.1371/journal.pone.0120870

Geurts, L. S., Cooke, J. R. H., van Bergen, R. S., & Jehee, J. F. M. (2022). Subjective confidence reflects representation of Bayesian probability in cortex. *Nature Human Behaviour*, *6*(2), 294–305. https://doi.org/10.1038/s41562-021-01247-w

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don’t know. *Proceedings of the National Academy of Sciences*, *113*(13), 3492–3496. https://doi.org/10.1073/pnas.1515129113

Guggenmos, M. (2022). Reverse engineering of metacognition. *eLife*, *11*(e75420).

Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *eLife*, *5*, e13388. https://doi.org/10.7554/eLife.13388

Hampton, R. R., Zivin, A., & Murray, E. A. (2004). Rhesus monkeys (Macaca mulatta) discriminate between knowing and not knowing and collect information as needed before acting. *Animal Cognition*, *7*(4), 239–246. https://doi.org/10.1007/s10071-004-0215-1

Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1668), 20140169. https://doi.org/10.1098/rstb.2014.0169

Kok, P., Brouwer, G. J., Gerven, M. A. J. van, & Lange, F. P. de. (2013). Prior Expectations Bias Sensory Representations in Visual Cortex. *Journal of Neuroscience*, *33*(41), 16275–16284. https://doi.org/10.1523/JNEUROSCI.0742-13.2013

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Kubanek, J. (2018). *Neuromodulation with transcranial focused ultrasound*. https://doi.org/10.3171/2017.11.FOCUS17621

Lawson, R. P., Bisby, J., Nord, C. L., Burgess, N., & Rees, G. (2021). The Computational, Pharmacological, and Physiological Determinants of Sensory Learning under Uncertainty. *Current Biology*, *31*(1), 163-172.e4. https://doi.org/10.1016/j.cub.2020.10.043

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, *63*(8), 1279–1292. https://doi.org/10.3758/BF03194543

Lockwood, P. L., & Klein-Flügge, M. C. (2021). Computational modelling of social cognition and behaviour—A reinforcement learning primer. *Social Cognitive and Affective Neuroscience*, *16*(8), 761–771.

Mamassian, P. (2016). Visual Confidence. *Annual Review of Vision Science*, *2*(1), 459–481. https://doi.org/10.1146/annurev-vision-111815-114630

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Mazor, M., Dijkstra, N., & Fleming, S. M. (2022). Dissociating the neural correlates of subjective visibility from those of decision confidence. *Journal of Neuroscience*, *42*(12), 2562–2569. https://doi.org/10.1523/jneurosci.1220-21.2022

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian Probability: From Neural Origins to Behavior. *Neuron*, *88*(1), 78–92. https://doi.org/10.1016/j.neuron.2015.09.039

Myers, C. E., Interian, A., & Moustafa, A. A. (2022). A practical introduction to using the drift diffusion model of decision-making in cognitive psychology, neuroscience, and health sciences. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1039172

Nelson, T., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of Learning and Motivation* (Vol. 26, pp. 125–173).

Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision*, *14*(11). https://doi.org/10.1167/14.11.5

Parr, T., Benrimoh, D. A., Vincent, P., & Friston, K. J. (2018). Precision and False Perceptual Inference. *Frontiers in Integrative Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnint.2018.00039

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

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8), 13. https://doi.org/10.1167/15.8.13

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience*, *19*(3), 366–374. https://doi.org/10.1038/nn.4240

Powers, A. R., Kelley, M., & Corlett, P. R. (2016). Hallucinations as top-down effects on perception. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *1*(5), 393–400. https://doi.org/10.1016/j.bpsc.2016.04.003

Press, C., Kok, P., & Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences*, *24*(1), 13–24. https://doi.org/10.1016/j.tics.2019.11.003

Rausch, M., & Zehetleitner, M. (2016). Visibility Is Not Equivalent to Confidence in a Low Contrast Orientation Discrimination Task. *Frontiers in Psychology*, *7*. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00591

Recht, S., Li, C., Yang, Y., & Chiu, K. (2024). *Adaptive curiosity about metacognitive ability*. OSF. https://doi.org/10.31234/osf.io/mhnvt

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current Research and Theory*, 64–99.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-09075-3

Rushworth, M. F. S., & Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, *11*(4), 389–397. https://doi.org/10.1038/nn2066

Sallet, J., Mars, R. B., Noonan, M. P., Neubert, F.-X., Jbabdi, S., O'Reilly, J. X., Filippini, N., Thomas, A. G., & Rushworth, M. F. (2013). The Organization of Dorsal Frontal Cortex in Humans and Macaques. *Journal of Neuroscience*, *33*(30), 12255–12274. https://doi.org/10.1523/JNEUROSCI.5108-12.2013

Sanders, J. I., Hangya, B., & Kepecs, A. (2016). Signatures of a Statistical Computation in the Human Sense of Confidence. *Neuron*, *90*(3), 499–506. https://doi.org/10.1016/j.neuron.2016.03.025

Scott, A., & Gilbert, S. (2024). *Metacognition guides intention offloading and fulfilment of real-world plans*. https://doi.org/10.31234/osf.io/y46mq

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science*. https://www.science.org/doi/abs/10.1126/science.270.5234.303

Shea, N. (2012). Reward Prediction Error Signals are Meta-Representational. *Nous (Detroit, Mich.)*, *48*(2), 314–341. https://doi.org/10.1111/j.1468-0068.2012.00863.x

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, *18*(4), 186–193. https://doi.org/10.1016/j.tics.2014.01.006

Shekhar, M., & Rahnev, D. (2018). Distinguishing the Roles of Dorsolateral and Anterior PFC in Visual Metacognition. *The Journal of Neuroscience*, *38*(22), 5078–5087. https://doi.org/10.1523/JNEUROSCI.3484-17.2018

Sheppard, J. P., Raposo, D., & Churchland, A. K. (2013). Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans. *Journal of Vision*, *13*(6), 4. https://doi.org/10.1167/13.6.4

Sherman, M. T., & Seth, A. K. (2021). *Effects of expected task difficulty on metacognitive confidence and multitasking*. https://doi.org/10.31234/osf.io/3gfp2

Sherman, M. T., Seth, A. K., Barrett, A. B., & Kanai, R. (2015). Prior expectations facilitate metacognition for perceptual decision. *Consciousness and Cognition*, *35*, 53–65. https://doi.org/10.1016/j.concog.2015.04.015

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences*, *13*(8), 334–340. https://doi.org/10.1016/j.tics.2009.05.001

Skewes, J., Frith, C., & Overgaard, M. (2021). Awareness and confidence in perceptual decision-making. *Brain Multiphysics*, *2*, 100030. https://doi.org/10.1016/j.brain.2021.100030

Sohoglu, E., & Davis, M. H. (2020, October 29). *Rapid computations of spectrotemporal prediction error support perception of degraded speech*. OSF. https://osf.io/b2jpt/

Van Marcke, H., Le Denmat, P., Verguts, T., & Desender, K. (2022). Manipulating prior beliefs causally induces under- and overconfidence. *bioRxiv*. https://doi.org/10.1101/2022.03.01.482511

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310–1321. https://doi.org/10.1098/rstb.2011.0416

Yon, D. (2021). Prediction and learning: Understanding uncertainty. *Current Biology*, *31*(1), R23–R25. https://doi.org/10.1016/j.cub.2020.10.052

Yon, D., de Lange, F. P., & Press, C. (2019). The predictive brain as a stubborn scientist. *Trends in Cognitive Sciences*, *23*(1), 6–8. https://doi.org/10.1016/j.tics.2018.10.007

Yon, D., & Frith, C. D. (2021). Precision and the Bayesian brain. *Current Biology*, *31*(17), R1026–R1032. https://doi.org/10.1016/j.cub.2021.07.044

Zoefel, B., Allard, I., Anil, M., & Davis, M. H. (2020). Perception of Rhythmic Speech Is Modulated by Focal Bilateral Transcranial Alternating Current Stimulation. *Journal of Cognitive Neuroscience*, *32*(2), 226–240. https://doi.org/10.1162/jocn_a_01490