Trinity College Dublin

The University of Dublin

Expectation in Sensory Processing and Perceptual Decision-Making

Kevin Walsh

School of Psychology

Trinity College

The University of Dublin

A dissertation submitted for the degree of Doctor of Philosophy to the School of Psychology at the University of Dublin, Trinity College

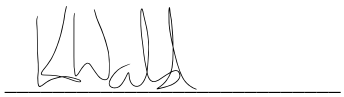
February 2022

Declaration

I, Kevin Walsh, declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).



Kevin Walsh
February 16th 2022

Summary

Sensory events cannot be directly translated into veridical percepts of their underlying causes because sensory information is inherently ambiguous and noisy. Indeed, any given pattern of retinal stimulation could arise from many arrangements of the sensory environment. The underspecification of sensory input appears insurmountable, but the brain must constrain this uncertainty to construct a useful perceptual experience. One solution to this problem is to exploit prior experiences of regularities in the sensory environment to generate expectations and narrow the pool of candidate hypotheses about the state of the world. If experience dictates that umbrellas are more probable occupants of hallways than pythons, the perceptual decision about the object lurking in the shadows can be informed by that expectation. No perceptual decision is made in the absence of any prior information, whether it is information about immediate contextual contingencies (e.g. pythons are not normally found in this hallway) or the long-term statistics of the sensory world (e.g. cardinal orientations are more prevalent than oblique ones). In recent years, there has been renewed interest in how prior experience might influence the conversion of sensory stimulation into perceptual experience. A growing literature suggests that the exploitation of priors is a fundamental feature of sensory processing, but the details of how exactly this is achieved is the subject of ongoing debate.

Chapter 1 offers an introduction to this debate and provides a review of the neurophysiological evidence for a contentious model at the centre of the controversy, called predictive processing, which claims that perceptual experience is the product of hierarchical inference.

According to predictive processing, the contents of perception will increasingly be shaped by expectation as the ambiguity of the sensory input increases. Chapter 2 describes an investigation of the predictive processing account of binocular rivalry, a phenomenon where perception oscillates between incompatible images shown separately to each eye. It was hypothesised that perceptual experience in the initial stages of rivalry could be biased by instilling an expectation that one of the rivalrous stimuli was more probable than the other. Prior experience did bias rivalry dynamics, but it was the unexpected stimulus that dominated perception in early rivalry. Explanations for this observation in terms of orientation-selective adaptation and the implications for the predictive processing account are discussed.

One of the divergent claims of predictive processing is that even the very earliest stages of sensory processing consist of comparisons between sensory expectations and sensory input. The experiment described in Chapter 3 assessed this contention by exposing participants to predictable sequences of contrast changes and measuring sensory responses to the violation of those sequences using an electrophysiological index of sensory processing, the steady-state visual evoked potential (SSVEP). Although there was evidence that the violation of expectation provoked a sensory response at a later stage of visual processing, there was no difference in the SSVEP response to expected and unexpected stimuli.

The thesis is composed of two halves; the first addresses some of the contentions of predictive processing and the second examines the mechanisms involved in prior-informed perceptual decisions. Chapter 4 provides an overview of the perceptual decision-making literature and describes the structure of a prominent model of the decision-making process, the Drift Diffusion Model.

According to sequential sampling models, perceptual decisions are the product of the integration of samples of sensory information over time until a threshold is reached. Extended evidence accumulation is an effective strategy when choices are based on physical evidence that is noisy or rapidly varying, but it is less obvious that this approach is beneficial when the evidence has little noise. Chapter 5 presents evidence that participants also accumulate evidence in perceptual decisions about low-noise stimuli. Tracing the response of electrophysiological decision signals to discrete pulses of sensory evidence provided estimates of the longevity of sensory samples in the decision process. Behavioural indices of the decision process were also used to assess the duration of this 'pulse effect'. In each case, the estimates far outlasted the proposals of decision models with little or no evidence integration.

The Drift Diffusion Model offers two mechanisms for the implementation of expectation in the decision process: a starting point bias and a drift bias. The vast majority of modelling studies have endorsed a starting point bias account with no drift bias, but recent evidence suggests that subtle decision adjustments may go undetected in traditional modelling approaches. Chapter 6 describes an investigation of the role of expectation in perceptual decision-making designed to detect signatures of these decision adjustments in neurophysiological data from three key stages of the sensorimotor hierarchy: sensory encoding, evidence accumulation, and response preparation. The study provided evidence that prior expectations lead to a starting point bias instantiated in biased response preparation, however several observations also suggested that prior-informed decisions are shaped by the complex interplay of multiple adjustments in the decision process. This will be further investigated in future modelling work.

Chapter 7 presents a general discussion of the thesis.

Acknowledgements

I am not exactly known for my sentimentality, but I would like to acknowledge the critical contributions of several colleagues and mentors to this project.

To Redmond, thank you for giving me the opportunity to pursue a PhD in the first place. I am certain that very few PhD students receive as much freedom and encouragement to pursue their ideas, and fewer still are offered as much time and support from their supervisors as I have been given over the last four and half years. Your excitement when things went well, reassurance when they didn't, and constant guidance throughout kept me on track in a process where it is easy to get lost. You have poured countless hours into getting me to this stage, but your patience and enthusiasm never changed and that made an enormous difference. I was very fortunate to have you as my supervisor.

To Dave, thank you for taking me under your wing as I finished my degree, making sure I got the chance to explore my interest in academia, first suggesting I consider a PhD, ensuring I secured funding, and being a constant source of advice and encouragement as a supervisor throughout this project. I undoubtably would not be here if it was not for you opening several doors for me. Thank you for all of the hours of work, the constant support, and the trust you had in me from the start.

To my friends in the O'Connell lab that made sure I looked forward to coming into the office each day. Whether the conversations were the EEG equivalent of "I don't know what a tracker mortgage is", debates about whether a blurry photograph of a road looks more like it's from Madagascar or Réunion, an education in the history of the Oscars, a collective whinge about Trinity bureaucracy, caustic invective over a game of Exploding Kittens, lengthy exchanges composed only of quotes from sitcoms, or just quiet chats about how things were going, there was never a dull day. I'm very grateful for all of your kindness and support. From hours spent participating in my experiments and reading drafts, to rations delivered to my house, to little messages checking I'm getting on ok and asking if anything can be done to help, it has made a big difference, particularly in the last few months.

To Mam and Dad, for every variety of support and encouragement throughout the PhD. I was ready to pack in education after a week of play school, but you insisted, and now here we are. I hope you're proud of yourselves. Thank you for ensuring I had everything I needed every step of the way. To Simon and Ciara, thank you for the coffee deliveries, the lifts home from the office on rainy nights, your solemn assurance that the thesis is completely typo-free, giving up more than a few weekends to participate in experiments, and keeping my spirits up after a day in the trenches. To Elizabeth, thank you for keeping me company, especially at meal times. There's nothing quite like being cloistered away for months amidst a pandemic to make you realise how critical all of the easily overlooked contributions are.

Finally, I would like to thank the Irish Research Council for funding this project (GOIPG/2017/1093).

Table of Contents

Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Special Issue: *The Year in Cognitive Neuroscience*
REVIEW

Evaluating the neurophysiological evidence for predictive processing as a model of perception

Kevin S. Walsh,[1] David P. McGovern,[1,2] Andy Clark,[3,4] and Redmond G. O'Connell[1]

[1]Trinity College Institute of Neuroscience and School of Psychology, Trinity College Dublin, Dublin, Ireland. [2]School of Psychology, Dublin City University, Dublin, Ireland. [3]Department of Philosophy, University of Sussex, Brighton, UK. [4]Department of Informatics, University of Sussex, Brighton, UK

Address for correspondence: Kevin S. Walsh, Trinity College Institute of Neuroscience and School of Psychology, Rm. 3.23, Lloyd Building, Trinity College Dublin, College Green, Dublin D2, Ireland. Walshk11@tcd.ie

For many years, the dominant theoretical framework guiding research into the neural origins of perceptual experience has been provided by hierarchical feedforward models, in which sensory inputs are passed through a series of increasingly complex feature detectors. However, the long-standing orthodoxy of these accounts has recently been challenged by a radically different set of theories that contend that perception arises from a purely inferential process supported by two distinct classes of neurons: those that transmit predictions about sensory states and those that signal sensory information that deviates from those predictions. Although these predictive processing (PP) models have become increasingly influential in cognitive neuroscience, they are also criticized for lacking the empirical support to justify their status. This limited evidence base partly reflects the considerable methodological challenges that are presented when trying to test the unique predictions of these models. However, a confluence of technological and theoretical advances has prompted a recent surge in human and nonhuman neurophysiological research seeking to fill this empirical gap. Here, we will review this new research and evaluate the degree to which its findings support the key claims of PP.

Keywords: predictive processing; perception; neurophysiology; perceptual inference; predictive coding

Introduction

Our immersion in a seamless and coherent perceptual experience veils the reality that it must be assembled from a sea of noisy and ambiguous sensory information.[1–3] It was Helmholtz who first proposed that perception does not obediently reflect the sensory inputs that undergird it, as exposed in myriad perceptual illusions,[4–6] but arises from an inferential process in which stimuli are interpreted in light of our past experiences. Quite how this is achieved in the brain is still hotly debated.[7–11] Thus far, the dominant framework guiding empirical research has been provided by feedforward models, which contend that our perceptual experience is assembled by a series of spatiotemporal filters that extract increasingly complex stimulus features as sensory information ascends through the cortical hierarchy.[12–14] These traditional models account for the influence of prior knowledge on perception by invoking a variety of processing strategies and computational heuristics[15,16] (e.g., the inhibition of short-range by long-range motion signals to solve the correspondence problem) that are tailored for particular sensory features and contexts. The success of feedforward models in accommodating neurophysiological and anatomical observations has established them as the orthodox perspective on sensory processing. In recent years, however, a growing contingent of scholars has entertained a radical new framework that casts perception as an entirely inferential process in which predictions about the outside world shape information processing at all levels of the cortical hierarchy.[1,4,9,17–19]

Predictive processing (PP) claims that the brain confronts the inherent ambiguity in sensory input by assembling "generative models" of the causes underlying sensory events. Generative models yield

doi: 10.1111/nyas.14321

Ann. N.Y. Acad. Sci. 1464 (2020) 242–268 © 2020 The Authors. *Annals of the New York Academy of Sciences* published by Wiley Periodicals, Inc. on behalf of New York Academy of Sciences

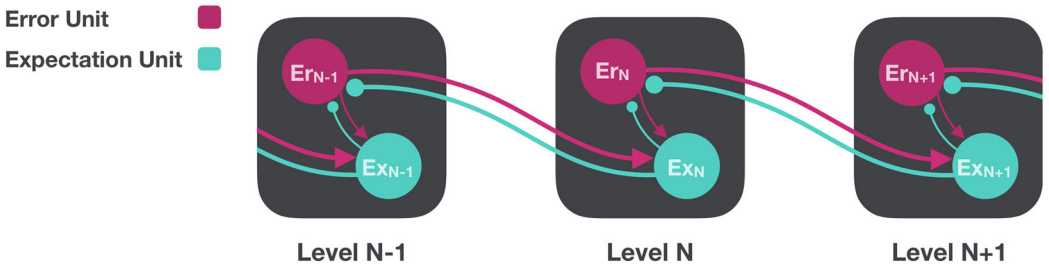


Figure 1. **A simplified hierarchical PP model based on the circuit proposed by Rao and Ballard.[18] Here, each level is composed of a population of reciprocally connected expectation units (Ex) and error units (Er). Expectation units signal the expected pattern of activity at the preceding level, given the current perceptual hypotheses. Where discrepancies emerge between these sensory expectations and sensory-driven input, prediction errors are fed forward to the subsequent level to inform the revisions of prediction necessary to minimize prediction error.**

predictions about the pattern of sensory input that would be expected if the model's estimate of the cause was correct. According to the dominant neural process account of PP, known as predictive coding,[18] these predictions are sent cascading down the processing hierarchy, suppressing congruent incoming sensory signals, such that only the residual, unexplained components of sensory information remain to be fed forward to higher levels in the form of "prediction error." The brain's generative models continuously exploit these error signals to revise the probability assigned to perceptual hypotheses as this iterative process plays out across all levels of the processing hierarchy, until the network converges on a consistent representation of sensory causes. From this perspective, perception is the process of identifying the perceptual hypothesis that best predicts sensory input and hence, minimizes prediction error.[9,20]

PP claims that this process is neurophysiologically instantiated in an inferential hierarchy composed of two functionally distinct neural subpopulations: expectation units that communicate expected sensory states downward and laterally within the processing hierarchy, and error units that feed prediction error signals upward and laterally[9] (Fig. 1). Importantly, prediction errors are not regarded as general surprise or arousal signals but rather, the source, connectivity, and stimulus preferences of an error unit imbue its output with specific information about the nature of the mismatch between predicted and actual input.[21] Thus, where traditional models propose that forward signals emanating from primary visual cortex (V1) directly reflect stimulus orientation, PP proposes that they exclusively encode deviations from an expected orientation. Equivalently, where traditional models propose that inferotemporal cortex (IT) detects object identities, PP proposes that its error units signal only unpredicted object identities. In other words, PP hinges on an inversion of classical feedforward accounts whereby it is the descending signals that provide representations of the external world while forward signals provide the feedback that modifies those representations. Thus, PP has at its core a process of "predicting the present," in which top-down flows attempt to match incoming sensory stimulations, but they achieve this goal using information that spans many windows of space and time[a].

To be effective, a PP system cannot treat all prediction error signals equally: sensory data can fail to conform to expectations because those expectations are incorrect or because the data are noisy or unreliable.[21,23] According to PP, the brain addresses this by adjusting the influence of particular sources of prediction error on perceptual hypotheses according to an estimate of the relative reliability associated with sensory evidence (i.e., sensory precision) and with higher-level representations (i.e., prior precision).[9,24,25] The "precision-weighting" assigned to prediction error is encoded in the synaptic gain applied to the associated error units, such that high-precision errors exert greater influence in commanding revisions of perceptual hypotheses, while current predictions are more stubborn in the face of low-precision

[a]Recent work in the area also looks at even longer windows, selecting policies that aim to reduce expected future prediction error (see Ref. 22).

errors.[26] Hence, precision-weighting provides a mechanism through which the brain can promote the influence of error channels that will provide the most valuable information in revising sensory expectations (see "Precision and attention" section).

A compelling feature of PP models is that they specify a single mechanism that naturally accounts for a range of prominent perceptual and neural phenomena ranging from end-stopping[18] to repetition suppression (RS) and error responses,[27] attentional modulations of sensory signals,[28] bistable perception,[29] and motion illusions.[30,31] But perhaps the most seductive aspect of these models is their apparent potential to provide a unified theory of the mind, with several theorists suggesting that hierarchical inference is the fundamental mechanism underlying all neural computation[1,9,17,19,21,32] (see Ref. 23 for a critical review). However, an oft-repeated criticism of PP is that it simply lacks the empirical foundation to undergird these grand claims[10,33–36] (see commentaries on Ref. 17). In reality, many of the perceptual and neurophysiological phenomena that are regularly highlighted as evidence in favor of PP can also be accommodated within traditional models by invoking additional mechanisms (e.g., neural adaptation in the case of RS). Most of PP's unique predictions relate to fine-grained neuronal and circuit-level phenomena that require carefully tailored behavioral paradigms and sensitive neural assays. A variety of PP-consistent models[b] have been proposed,[2,9,18,37–41] but despite this heterogeneity, there are a number of shared, canonical features that clearly dissociate PP from traditional models of perception. To date, neurophysiological investigations of PP have largely centered around testing four key hypotheses:

1. Error-signaling neural responses to sensory stimuli should scale inversely with expectation.
2. Top-down signals represent sensory prediction.
3. At each level of the cortical hierarchy there are two functionally distinct neural subpopulations representing predictions and prediction errors.
4. Prediction error minimization is achieved through reciprocal exchange of error and prediction signals across levels—a process known as "hierarchical inference."

In the last 5 years, methodological advances and the increasing reach and influence of PP in the neurosciences have prompted a significant surge in the number of neurophysiological investigations seeking to definitively test these claims. Here, we seek to offer a balanced overview and critical analysis of the current state of the evidence, highlighting some of the key studies that exemplify recent progress in the field (see Figs. 2–6).

Hypothesis 1: Error-signaling neural responses to sensory stimuli should scale inversely with expectation

The vast majority of neurophysiological investigations of PP have centered on the proposal that error-signaling neural responses to sensory stimuli should scale inversely with expectation. Much of what has been considered the core evidence for PP comes from experiments that test for three basic phenomena: vigorous neural responses to the omission of an expected stimulus (the "omission response"), the suppression of neural activity following repeated stimulus presentations ("repetition suppression"), and the suppression of neural activity following stimuli that are expected on the basis of statistical regularities or prior cues ("expectation suppression"). As we will see in this section, research in this domain has relied heavily on noninvasive recording methods that provide population-level measures of brain activity. This is potentially problematic because, in many instances, PP makes opposite predictions regarding the impact of a given experimental manipulation on expectation versus error unit activity and it is not clear how these distinct modulations should manifest in global neural responses. Although this is not a reason to dismiss the neural phenomena reported in this section, it is an important caveat when relating them to PP.

[b]Note that there are other biologically plausible accounts of predictive processing (e.g., belief propagation and variational message passing), which many take to be synonymous with active inference, although they would not necessarily involve prediction error minimization, but here we focus on predictive coding because it has been by far the dominant account considered by the extant neurophysiological literature.

Omission responses

Perhaps the most compelling example of the putative prediction error signal is the robust neural response elicited by the *omission* of an expected stimulus[42–48] (see Ref. 42 for a review). Importantly, rather than being generic surprise responses, several studies have identified omission signals that are feature-selective. For example, omission responses in primary visual cortex have been shown to be retinotopically specific and exhibit comparable feature specificity with the responses that are evoked when the stimulus is presented.[49] Although PP can comfortably accommodate prediction error signals being elicited in the absence of any sensory input, such omission responses do not appear easy to reconcile with feedforward models that are devoid of any predictive influence on sensory processing[41] (but see Ref. 50). However, PP's interpretation of the omission response also suffers from some ambiguities. For example, if prediction is simply subtracted from bottom-up activity, the prediction error response to the omission would require negative firing rates (see "Evidence of prediction in sensory processing" section). This has led some to suggest that omission responses may predominately reflect the activity of expectation units rather than error units.[37] Indeed, it might be hypothesized that the omission response is composed of activity in expectation units representing the sensory prediction of the absent stimulus[51–53] and the error response to the mismatch with those expectations. Thus, while the observation of neural responses to unexpected stimulus omissions provides evidence for the role of prediction in sensory processing, further research is required to work out their precise neurophysiological origins.

Repetition suppression

RS has been consistently reported across a wide range of methodologies, sensory modalities, stimulus properties, and time scales.[54,55] Although it accords neatly with PP,[27] an equally plausible interpretation of RS is that it arises from low-level changes in the responsivity of stimulus-selective neurons (i.e., neural adaptation).[56] Thus, the key test for PP models is not whether RS occurs, but whether its occurrence can be directly attributed to expectation. An influential early study by Summerfield *et al.*[57] sought to control for the effects of adaptation by comparing blood oxygen level—dependent (BOLD) activity in the fusiform face area (FFA) of human participants in response to stimulus repetitions that were expected versus unexpected, and found that expected repetitions yielded substantially smaller BOLD responses. Numerous subsequent functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and electroencephalography (EEG) studies have reported modulations of RS by the probability of a repetition[45,46,58–67] (i.e., expectation) (but see Refs. 68–70). Also consistent with PP, and seemingly at odds with neural adaptation accounts, is evidence that unexpected repetitions evoke greater neural responses than frequent alternations.[37] However, as we will discuss in more detail below (Hypothesis 4), PP asserts that expectations due to stimulus repetition and those due to knowledge of stimulus probability emanate from distinct processing levels. Grotheer and Kovacs[61] implemented orthogonal manipulations of face repetition and repetition probability, noting that studies potentially confound the effects of repetition and expectation on sensory signals by manipulating the relative probability of repetitions and alternations across blocks. They observed independent, additive effects of each in FFA, the occipital face area, and the lateral occipital complex (LOC).

Attempting to replicate the findings of Summerfield *et al.*[57] with single-cell recordings in macaque IT, Kaliukhovich and Vogels[71] reported that responses to deviant stimuli were indistinguishable from responses to infrequent, but conditionally probable, stimuli. However, it is unclear to what extent this discrepant result reflects certain methodological differences. First, Kaliukhovich and Vogels used stimuli that were unfamiliar to the monkeys, but some human studies indicate that RS expectation effects are most pronounced for stimuli that are highly familiar to the observer[59,61] (e.g., letters or faces; but see Ref. 63). Second, the task demanded only passive fixation rather than attentional engagement, which has been shown to strongly affect the modulation of RS by expectation.[60] Third, Kaliukhovich and Vogels recorded from IT, while Summerfield *et al.* measured activity in FFA. To address these issues, Vinken *et al.*[72] recorded neural activity in response to face stimuli in the middle lateral face patch (ML), a macaque homolog of FFA. Once again, they failed to observe any evidence of an effect

of repetition probability on RS in single-unit or multiunit activity, even when the repetitions were task-relevant and repetition probability was found to affect the monkeys' behavior. However, Vinken *et al.* measured responses from ML, which is characterized as having view specificity rather than identity specificity.[73] As alternate stimuli had a different identity, but a similar viewpoint, these stimuli may not have optimally differentiated "expected" and "unexpected" responses. Importantly, the results from both Vinken *et al.* and Kaliukhovich and Vogels were presented as single population analyses, which would also obscure any functional heterogeneity among individual neurons representing error units or expectation units. Thus, it remains an open question if the effects of RS are a consequence of perceptual expectations,[27,55] but the fact that PP can accommodate RS as a purely low-level phenomenon[27] suggests that RS may not represent a useful vehicle for definitively adjudicating between PP- and feedforward/adaptation-based accounts.

Expectation suppression

A substantial body of work has also sought to test PP predictions pertaining to expectation suppression (ES) using a variety of paradigms (e.g., predictive cues, paired associations, and predictable stimulus sequences). As with RS, ES has been consistently reported in human brain imaging studies across sensory modalities and noninvasive neural recording techniques[45,46,74–81] (but see Ref. 82) and a number of human studies have demonstrated that the magnitude of stimulus-evoked responses is inversely proportional to the degree to which the stimulus was expected[75,76,83,84] (see Refs. 47 and 85 for corresponding evidence from neuronal recordings in rodents and monkeys, respectively). A key observation from this work has been that the observed neural modulations are not generic to broad categories of stimuli, but specific to individually predicted exemplars. For instance, Pajani *et al.*[70] observed ES in FFA BOLD activity in response to specifically predicted facial identities relative to nonpredicted faces.

Although many of the above studies could not segregate the distinct contributions of error and expectation units to the observed global neural response modulations, some studies have applied alternative experimental designs and methodologies that were tailored to do so. Egner *et al.*[76] examined BOLD responses to face and house stimuli that were either expected or unexpected, and observed effects consistent with ES in FFA. The authors also made the surprising observation that, despite its well-established role in face processing,[86] FFA responses to faces and houses were statistically indistinguishable when those stimuli were expected. The full pattern of effects was best explained by a PP model in which BOLD responses were dominated by strong prediction-driven, face-selective expectation unit activity and weak face-selective error unit activity on both trial types (since houses would evoke minimal error signals from face-selective cells). Another fMRI investigation by de Gardelle *et al.*[87] attempted to isolate activity attributable to each subpopulation when examining RS. The authors hypothesized that activity in expectation units should be enhanced as repetitions generate increasingly precise sensory expectations, while error units should exhibit RS as these predictions eradicate prediction error. Consistent with these predictions, two segregated clusters of FFA voxels were found to exhibit either RS or repetition enhancement. Although the authors demonstrated that the classification of each voxel was stable across measurements, the inevitable intermingling of the proposed expectation and error unit populations within each 3-mm cubic voxel is a limitation of this approach. For example, Auksztulewicz and Friston[27] point out that increased expectation-linked activation of a given voxel cannot be reliably attributed to expectation unit activity because the same trend could arise if prediction error units in that voxel had been assigned increased precision-weighting as the task statistics are learned.

A series of single- and multiunit recording studies have also reported evidence of ES in macaque IT neurons.[85,88–92] These effects were sensitive to transitional statistics (e.g., the transition from Stimulus A to Stimulus B being more probable than from B to A),[90,91] persisted after controlling for RS,[88,89] and emerged at early latencies ($\sim$150 ms).[88,90–92] For example, Bell *et al.*[88] recorded single-unit activity from IT neurons while monkeys performed a delayed match-to-sample task in which they indicated which of two stimuli (a face or a fruit) best matched a previously displayed cue stimulus that was degraded by noise. They measured the effect of stimulus probability on IT responses by manipulating the relative probability of the cue

being a face or a fruit, which changed unpredictably over the course of the recording session. The results showed a reduced population response to expected faces in face-selective IT neurons, which Bell *et al.* interpreted as evidence that IT neurons encode prediction errors and long-term probabilistic information, in line with PP. Although many of these invasive ES studies are subject to the same criticism of failing to segregate expectation unit and error unit activity as the studies from the RS literature, the analyses of Bell *et al.*[88] did identify two distinct populations underlying the recorded activity (see Hypothesis 3), with the ES effect evident in the putative error unit subpopulation specifically.

However, Vinken and Vogels[93] showed that this reduced response in IT neurons could be reproduced in a simulated population of neurons through stimulus-specific neural adaptation without recourse to PP mechanisms. The authors simulated a population of neurons displaying stimulus-selective responses that decreased with each presentation of a preferred stimulus and recovered between presentations, allowing suppression to build up over time. This build-up of suppression over time ultimately accounted for the putative ES effects, indicating that they could be explained by passive adaptation mechanisms;[93] although there has been subsequent debate as to whether the long time window over which adaptation operates in the simulations is realistic given previous empirical observations.[94] The picture is further complicated by the fact that a number of direct recording studies have produced the opposite pattern of results to *Bell et al.*, with unexpected or random stimuli eliciting reduced responses relative to expected or neutral stimuli in macaque IT[95,96] and rodent V1[52] (see Ref. 97 for corresponding results in a human study). There is also conflicting evidence regarding whether IT neurons are sensitive to statistical regularities present in nonadjacent stimuli from a sequence of images[91] or whether these suppression effects are determined by the immediately preceding stimulus.[89,96]

Precision and attention

One promising avenue for disentangling PP and adaptation-based accounts arises from PP's specification that the magnitude of expectation effects on neural activity depends on the precision-weighting applied to particular error signals. To efficiently minimize prediction error, the system adjusts the precision-weighting assigned to prediction error according to the estimated reliability of sensory information. Hence, the amplitude of error responses will reflect both differences in top-down predictions and differences in predicted precision. Indeed, as mentioned above, many studies have reported that stimulus-evoked responses are inversely proportional to the degree to which the stimulus was expected,[47,75,76,83–85,98] although this effect is not universally observed.[79,82]

In PP, stimulus salience and attention are considered emergent properties of this precision-weighting mechanism.[9,28] Here, attending to a stimulus feature or to a part of the visual field is equivalent to predicting high-precision information from the associated error units and hence upweighting their error signals' influence in subsequent revisions of perceptual hypotheses. This is consistent with the results of numerous studies that report larger neural responses to attended compared with unattended stimuli.[99–101] Although this finding is easily reconciled with the traditional feedforward account of attention, PP diverges from this account in proposing an interaction between attention and expectation. Characterizing the nature of this interaction has proven difficult because the experimental designs found in much of the earlier research confounded attention and expectation[3] and the heterogeneity of results from subsequent efforts to independently manipulate these variables precludes any conclusive resolution. For example, the prominent interaction model suggests that the error response to expected stimuli should be amplified for attended, but suppressed for unattended, stimuli, while the reverse should be true for unexpected stimuli.[102,103] Although several reports are consistent with this pattern,[102,104,105] others have found that ES is abolished in the absence of attention,[60,106,107] or that attention provides an equivalent boost to predicted and unpredicted stimuli.[77] Hsu *et al.*[108] argue that the source of these conflicting results may be the distinction between unpredicted and mispredicted. They found that the response suppression evident in comparisons of predicted and mispredicted stimuli showed no modulation by attention. However, when comparing responses with predicted and unpredicted stimuli, attention reversed ES.

Hypothesis 1: Summary and conclusions

Although many studies report the expectation-related neural modulations that are predicted by PP across the full range of neurophysiological recording techniques, a significant contingent failed to replicate these effects. For instance, there is a clear disparity in the frequency with which expectation-related modulations are detected using global measures compared with direct neuronal recordings. The uncertainty surrounding the relationship between the neural activity preferentially recorded with invasive versus noninvasive assays[54] also hinders efforts to integrate observations across the literature. In addition, this summary of the recent literature highlights considerable variation in the paradigms that have been employed. For instance, while some studies instill expectations by allowing subjects to learn stimulus probabilities through exposure (e.g., Ref. 66), others train subjects for days (e.g., Ref. 92); some manipulate baseline probabilities (e.g., Ref. 63), others manipulate transitional probabilities (e.g., Ref. 90); some engage attention (e.g., Ref. 45), others involve passive viewing (e.g., Ref. 71); some inform subjects about the manipulation of conditional probabilities (e.g., Ref. 81), others do not (e.g., Ref. 57). This methodological heterogeneity precludes direct comparisons across studies since PP makes divergent predictions regarding the properties of prediction error given different permutations of these variables.

Another source of uncertainty in interpreting these effects arises from the fact that there is no consensus about precisely which components of neural activity are being suppressed by expectation (see Ref. 4 for a recent review). Although some studies have reported that expected stimulus representations are dampened,[69,79,90,92,105,109–111] others have found that this global suppression is attributable to the suppression of neurons tuned away from the expected stimulus feature, while the representation of the expected stimulus is actually sharpened[c].[78,88,112] Interestingly, Marques *et al.*[113] recently found that while backward inputs from rodent lateromedial visual area (LM) to V1 were retinotopically matched on average, they also divaricated widely beyond the target receptive field, relaying distal visual information. In fact, half of V1 inputs from LM had receptive fields displaced more than 24° from their target in V1 and these axons increasingly targeted cells with the opposite tuning profile as the retinotopic distance between the LM and V1 cells increased, betraying a complex circuitry that would be difficult to uncover using global measures of brain activity. Additionally, as ES is typically calculated by comparing expected and unexpected stimulus responses, it is often unclear if it represents genuine suppression, suppression relative to the enhanced response to deviant stimuli, or both.[55,88,96] For example, Kaposvari *et al.*[89] provided evidence of both: an early transient suppression of the neural response to expected stimuli and later sustained enhanced activity in response to unexpected stimuli, reminiscent of a prediction error response.

A central criticism of many of the studies addressing Hypothesis 1 is the tendency to attribute fluctuations in feature-selective global neural responses to error signals while overlooking the fact that, according to PP, expectation units are concurrently active and undergo distinct modulations[d].[34] Within a PP framework, global measures of neural activity, such as MEG/EEG and fMRI, will necessarily reflect a summation of activity from each subpopulation and the magnitudes of their relative contributions are not easily derived *a priori*. This is particularly important because PP makes opposite predictions depending on whether activity represents the transmission of sensory predictions[49,82,97,114,115] (enhanced activity) or the integration of these predictions with bottom-up activity in prediction error units (reduced activity).[45,46,74–81] Although this problem is not unique to PP, the indefinite profile of neural activity captured in BOLD activity[116,117] and MEG/EEG signals[118] hampers efforts to provide dispositive evidence of PP in humans.[34]

However, PP does make a number of more detailed specifications regarding the architecture of the inferential circuitry that recent work has begun

[c] When interpreting these results, it should be noted that these decoding analyses of global BOLD activity or single-unit responses did not segregate activity from the proposed error unit and expectation unit populations.

[d] It should be noted that some have suggested that EEG recordings will preferentially sample the activity of error units due to their association with superficial pyramidal cells (e.g., Refs. 9 and 24; see Hypothesis 3).

to address. First, PP proposes that predictive signals carried in descending cortical connections are responsible for the suppression of prediction error responses to expected stimuli (Hypothesis 2). Second, PP suggests that these signals originate in two functionally distinct neural subpopulations (error units and expectation units) and that their activity should reflect expectations in very different ways (Hypothesis 3). Third, PP proposes an inferential hierarchy whereby expectation modulations are not uniform across the cortex but are selectively applied at the processing levels to which the expectations pertain (Hypothesis 4). Thus, the particular neural populations and processing levels that are preferentially sampled by different recording techniques may be key to the study outcome. This may account for the fact that, in several instances, RS/ES effects that were observed with one noninvasive recording method were absent in another despite using otherwise identical paradigms.[53,78,119] This issue greatly complicates efforts to generate precise, empirically testable PP hypotheses and also reduces the scope for disconfirmatory results.[42] In recognition of this point, researchers are increasingly seeking to verify the other hypotheses of PP by attempting to isolate mechanisms underlying the generation of predictions and hierarchical inference, and it is to this literature that we now turn our attention.

Hypothesis 2: Top-down signals represent sensory predictions

A defining characteristic of PP is a cascade of descending predictions distributed to each level of the processing hierarchy to quell emerging prediction error.[17] Human and animal studies indicate that the brain is finely tuned to efficiently extract regularities from streams of sensory input[120,121] through passive sensory experience,[52,122] self-generated actions,[123,124] and exploration of the environment,[47,124] and that these expectations influence sensory processing.[45,74,78,125,126] Although many agree that feedback plays an essential role in myriad sensory processes,[127,128] the neural instantiation of these influences is more controversial. As outlined in the preceding section, the vast majority of studies examining PP have tended to focus on prediction error responses, but a number of recent studies have also sought to probe the nature of neural prediction itself.

Evidence of prediction in sensory processing

Devising experimental paradigms that can isolate stimulus-specific prediction signals from other coincident neural activity is far from straightforward. One creative approach has been to exploit the Kanisza illusion in which "Pac-Man" shapes are rotated to manipulate the perception of an illusory triangle (Fig. 3B) in order to probe for prediction responses in the absence of any bottom-up input. In a departure from much of the work described above, these predictions emerge from the interpretation of a currently viewed stimulus rather than being based on experimentally controlled stimulus probabilities. Employing the Kanisza paradigm, Kok and de Lange[129] observed increased BOLD activity in regions of V1 and V2 retinotopically mapped to the illusory triangle contours. However, it is not clear whether this activity should be interpreted as representing error unit activity arising from the absence of sensory input where triangle contours are expected or prediction unit activity reflecting the perceptual hypothesis itself.[34] Single-unit recording studies have indicated that neural responses to illusory contours reflect the descending influence of higher-level areas.[130,131] For example, while both macaque V1 and V2 neurons respond to the illusory contour of a Kanizsa triangle, V2 neurons consistently respond earlier than those in V1.[130] Several fMRI studies have also reported that patterns of neural activity evoked in early visual cortex following the omission of an expected stimulus resemble the activity evoked by the veridical stimulus[49,51,132] and this result has been replicated in recordings from rodent V1.[52] Muckli *et al.*[132] occluded a subregion of a visual scene and isolated neural activity in voxels retinotopically mapped to this area of the visual field (Fig. 6B). In line with the idea that predictive signals descend from higher-level processing areas, they observed that the classification performance of their decoding analysis was robust to 2° shifts in the visual scene, suggesting that this backward activity originates in neurons with larger receptive fields.

Elsewhere, several studies have offered evidence that sensory prediction may also rely on representations of stimulus trajectories leading into the future. For example, Bendixen *et al.*[43] played participants a series of pairs of identical tones and, in certain instances, omitted either the first or second tone in a pair. Omission responses were only observed

when the identity of an omitted tone was known in advance of its omission (i.e., the second tone of the pair was omitted; see also Ref. 44). Also consistent with this aspect of PP, enhanced anticipatory feature-selective activity preceding high-probability stimuli compared with low-probability stimuli has been reported in MEG recordings of visual cortex,[53] BOLD activity in FFA,[82] and single-unit recordings in macaque IT.[88] Bell *et al.*[88] and Kok *et al.*[53] showed that a representation of the expected stimulus could be decoded from neural activity before stimulus onset using multivariate analyses and forward-modeling, respectively. However, it should be noted that the prestimulus expectation effects observed by Trapp *et al.*[82] and Kok *et al.*[53] did not influence the subsequent stimulus-evoked response.

In another line of research, Van Kerkoerle *et al.*[133] trained monkeys to mentally retrace a briefly presented (150 ms) curve and recorded V1 activity. Following the disappearance of the curve stimulus, spiking activity persisted in superficial and deep layers of V1 in the same retinotopic locations as the presented contour, suggesting that V1 contains a persistent trace of recently presented stimuli. Interestingly, the activity was present up to 600 ms after stimulus offset and was temporarily erased by the presence of a visual mask, but reinstated on mask removal, demonstrating that the activity could not be explained through iconic memory mechanisms.[133] In a human fMRI study, Ekman *et al.*[51] repeatedly exposed participants to a dot stimulus rapidly moving across a screen. Following the exposure period, a flash of the dot stimulus at the starting position produced a time-compressed sequence of BOLD activity across the retinotopic locations of V1 corresponding to the previously observed stimulus trajectory. Importantly, this "preplay" activity was not elicited when the stimulus was flashed at the end point of the moving dot sequence (Fig. 2). This same phenomenon has previously been directly observed in the firing activity of an ensemble of V1 neurons in rodents[52,122] and in macaque V4.[134] Gavornik and Bear[52] point out that the temporal specificity of these sequence representations are not predicted by Hebbian plasticity. Chong *et al.*[135] measured V1 BOLD responses to an apparent motion illusion composed of spatially separated presentations of a rotating grating. V1 activity was found to contain an interpolated representation of an intermediate grating orientation along the illusory motion trajectory between the two grating presentation locations. Given that this intermediate grating was never presented to participants in the experiment, this result suggests that the brain reconstructs dynamic object features in V1, predicting current sensory input based on a representation of the stimulus trajectory.

Further evidence of neural prediction signals has been uncovered in research examining the substantial locomotor contributions to activity in rodent V1. These studies cleverly exploit a divergence between traditional models of sensory processing and PP by using self-motion as a proxy for motor-related sensory prediction.[41] Although traditional feedforward models do not predict a difference in neuronal responses in early visual areas to visual flow depending on whether it is self-generated or externally caused, PP suggests that error units will signal any deviation between sensory predictions and sensory input. Indeed, when mice run through a virtual environment, neural activity in V1 is significantly modulated by locomotor feedback carrying information about expected patterns of stimulation.[47,48,124,136] For example, the activity of anterior cingulate cortex (ACC) axons projecting to rodent V1 has been shown to convey predictions of upcoming grating stimuli[47] or visual flow[124] based on the mouse's movements. The V1 neurons targeted by these backward connections have been found to signal the mismatch between predicted and actual sensory input for their corresponding region of the visual field, characteristic of PP error units.[48,98] These predictions do not simply modulate visual responses but appear to drive the activity of V1 neurons even in the absence of bottom-up input.[48,136] Moreover, these axonal influences are found to be experience dependent.[47,123,124] For example, Leinweber *et al.*[124] found that this predictive feedback to V1 came to reflect a new visuomotor coupling when a mouse was trained in an inverted virtual environment. Overall, these studies are not consistent with models characterizing primary visual cortex as a passive feedforward filter.[128]

At the same time, demonstrations of top-down predictive influences on aspects of sensory processing do not necessarily provide definitive evidence for PP. For example, although Saleem *et al.*[136] identified a small number of neurons sensitive to mismatches between input and expectations

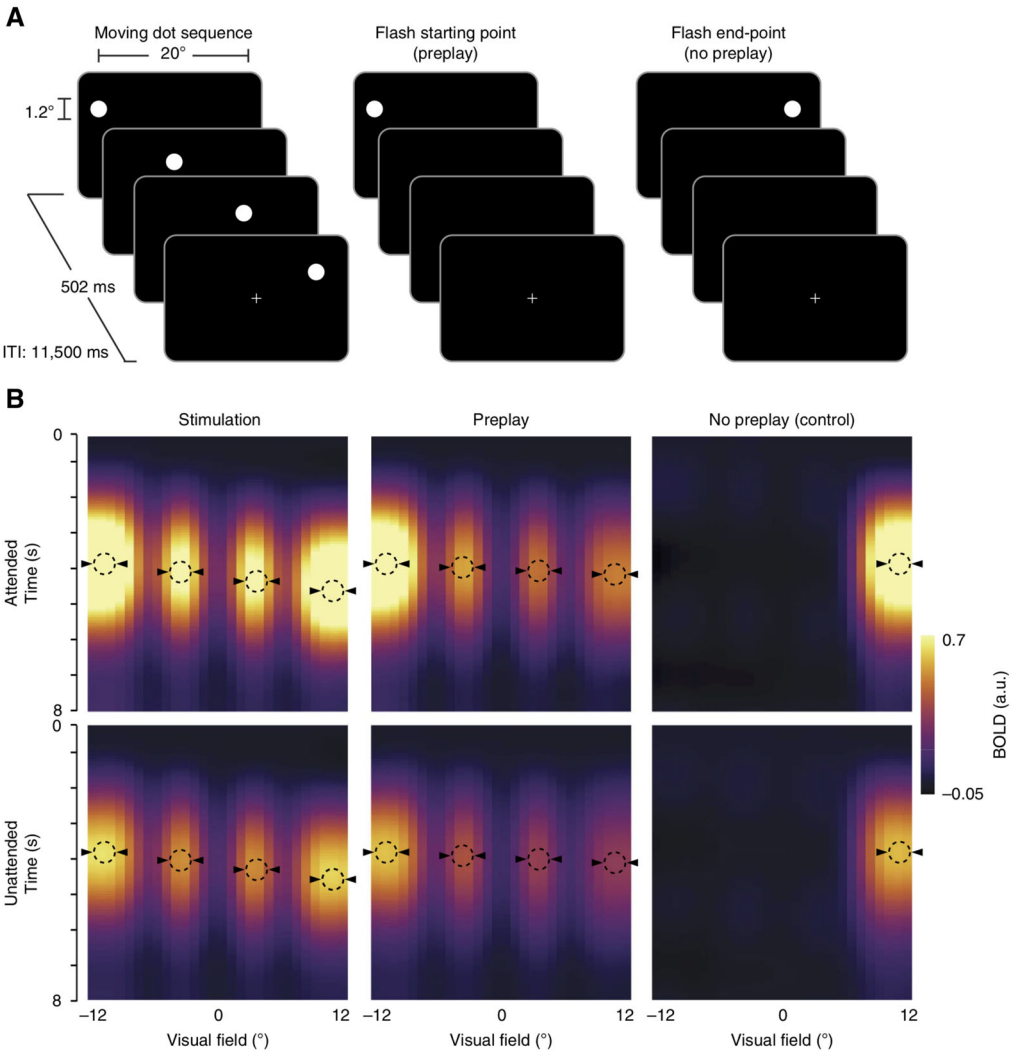


Figure 2. Task schematic and data from Ekman *et al.* (adapted from Ref. 51). (A) Human participants were repeatedly presented with a moving dot sequence for 4 minutes. Following this dot sequence, participants were then shown either the starting point of the sequence or the endpoint, which was briefly flashed on the screen. (B) The starting-point stimulus generated a sequence of BOLD activity across the retinotopic locations of V1 corresponding to the positions of the actual dot stimulus, which reconstructed the stimulus sequence in a time-compressed format (i.e., this activity unfolded more rapidly than the response to the actual stimuli). This "preplay" of the stimulus sequence was not elicited by the endpoint stimulus and was still observed in the absence of attention. The authors argued that the time-compressed format indicated that this represented automatic predictive activity and not surprise at the omitted stimuli. Enhanced activity in hMT indicated that this activity is fed back from higher-level regions.

in V1, they found that the feedback modulation of the wider V1 population was better accounted for by a positive linear weighted sum of sensory input and predicted input than by prediction error. In addition, PP mismatch responses are typically considered to be the product of a simple subtraction (error = actual input – predicted input), but this scheme would appear to necessitate negative firing rates for omitted stimuli or for weaker than expected sensory inputs. In fact, several rodent studies have observed just such a response to an expected but omitted stimulus.[47,48,98,123] Keller and Mrsic-Flogel[41] accommodated the representation of signed prediction errors by proposing that PP

is subserved by two classes of error unit: positive error units (excited by bottom-up sensory signals and inhibited by top-down prediction) and negative error units (with the opposite mapping). However, a recent study by Spratling[137] showed that simulations using an alternative PP model,[2] which uses division rather than subtraction to calculate error, better fit the neurophysiological data derived from these rodent studies than either Rao and Ballard's original model[18] or Keller and Mrsic-Flogel's revised model.[41]

The development of prediction

According to PP, the range of interpretations of sensory stimulation entertained by the perceptual system is constrained by prior experience. Some priors can be bestowed through phylogenetic development and hardwired into neural structure,[138] while others will emerge from persistent regularities in natural sensory experience and will likely be instantiated in the tuning properties of sensory cortex.[4,19,139] For example, the prevalence of cardinal orientations in the sensory environment is reflected in the more narrow tuning and overrepresentation of neurons selective for cardinal orientations in early visual areas.[140] Berkes *et al.*[141] found that spontaneous activity in ferret V1 becomes increasingly consistent in response to natural scenes over the course of development, indicating that the visual system converges on a response shaped by prior sensory experience.

Nonetheless, such consistently reinforced priors appear malleable in the face of new experience,[6,142,143] which suggests they may be represented both in the architecture of sensory cortex and in more dynamic, context-sensitive top-down influences. In this way, priors based on recent patterns of stimulation or current context can also influence sensory processing,[2,144–147] even when regularities are embedded in complex naturalistic stimuli.[139] For example, Li and Di Carlo[148] designed a paradigm that exploited the fact that object representations in IT are robust to changes in viewing angle. As monkeys freely viewed a monitor, an image of an object was presented 3° displaced from their retinal position. When the monkey spontaneously saccaded to the image, the identity of the object was alternated during saccade execution. Subpopulations of neurons that initially exhibited a postsaccadic response preference for the first object identity, came to incorporate both object identities equally after repeated exposure, which is consistent with the idea of a flexible predictive model capable of updating stimulus expectations to reflect sensory regularities.

Several rodent studies have provided evidence that predictive activity emerges over the course of exposure to statistical regularities in visual stimuli[47,52,124] (see also Ref. 88 for single unit recording in monkeys) and that this correlates with the magnitude of prediction error responses.[47,123] In line with PP's precision-weighting mechanism, there is also evidence from auditory studies in humans that the inherent predictability of a stimulus stream moderates this process of prediction generation[144,149,150] and subsequent ES.[151–153] For example, Southwell and Chait[154] found that the neural response to frequency outliers was enhanced when the deviant tones were presented in the context of a regular pattern compared with a random pattern even when the pattern was presented too rapidly to allow conscious detection of the regularity.

A particularly perplexing set of results in this literature comes from a series of studies reporting that the transition from random to regular sequences of auditory stimuli is accompanied by a sustained increase in neural activity measured with MEG, EEG, and fMRI,[144,154,155] in stark contrast to ES studies showing dampened neural responses to predictable stimuli.[78,81,156,157] One interpretation of this result is that the enhanced activity reflects the high precision-weighting afforded to low variance input.[155] However, since precision-weighting is applied to error units, this explanation of the sustained error signals rests upon the assumption that the increased precision more than compensates for the reduction in prediction error that would be expected over the course of exposure to a repeating, and therefore highly predictable, pattern of tones.

Finally, recent studies provide evidence of prediction error updating perceptual hypotheses to improve future predictions.[69,158] For example, Tang *et al.*[69] applied a forward encoding model to EEG data in order to measure the orientation selectivity of visual signals in response to pairs of oriented Gabors. Unexpected orientations led to increased orientation selectivity soon after stimulus presentation and this effect reemerged at later time points, consistent with an updating of sensory expectations.

Overall, this evidence is consistent with PP's contention that the brain constructs dynamic representations of regularities in the sensory stream and that these representations are intimately involved in subsequent sensory processing.[144,154,159]

Hypothesis 2: Summary and conclusions

Humans are capable of rapidly extracting regularities from the sensory environment[144,158,160,161] and there is strong evidence that the resultant expectations influence sensory processing.[154] It is a substantial achievement to begin to isolate this predictive activity in recordings of neural activity spanning human, monkey, and rodent research. Consistent with PP, neural activity believed to represent prediction appears to carry stimulus-specific information, which is heavily experience dependent, and interacts with bottom-up sensory input. However, there are many questions still to be answered. Although significant progress has been made,[4,42,162] the neurophysiological mechanisms responsible for the extraction of regularities and the generation of experience-based priors are not well understood. Although a range of candidate neural regions have been implicated in tracking a variety of sensory regularities and issuing predictions[29,37,40,53,55,75,120,144,158,162] (see Ref. 41 for a recent review), their precise relationships to the modulations observed in sensory processing are not yet established. Finally, it is not clear whether the formation of sensory predictions is a unitary neural process or an array of independent, task-tailored mechanisms.[161] Expectations can be formed in relation to stimulus timing, stimulus location, or stimulus content,[162] and electrophysiological evidence suggests that these kinds of predictions are instantiated by distinct neuromodulatory mechanisms, in dissociable networks and at different latencies.[44,162,163] Similarly, expectations can arise from arbitrary stimulus pairings,[90] predictive cues, or higher order regularities.[66] To what extent these forms of prediction rely on the same basic neural mechanisms remains to be determined.

Hypothesis 3: Each level of the cortical hierarchy houses two functionally distinct neural subpopulations representing predictions and prediction errors

A central postulate of PP is that prediction error minimization is coordinated between two functionally distinct prediction and error unit subpopulations, which propagate signals across different frequency bands and cortical layers.[9,40,42] However, it is only very recently that empirical research has directly addressed this hypothesis.

Functionally distinct units

A corollary of PP's specification of two distinct subpopulations is that neurons encoding expectations should not encode prediction error and vice versa. Indeed, studies of primate and rodent neuroanatomy indicate that forward and backward projections originate from separate cell populations,[164,165] as PP requires. As we have seen, some fMRI studies have attempted to account for these subpopulations in their analyses (e.g., Refs. 76 and 87), but ultimately, BOLD responses will not adequately segregate their output as each voxel would contain an uncertain mixture of error and expectation units. More compelling evidence of functionally distinct subpopulations has been acquired through examination of single-unit recordings. For example, Bell *et al.*[88] found no correlation between the activity of a subset of neurons that encoded stimulus probability (i.e., prediction) and a subset that encoded prediction errors in macaque IT. Also in line with the predictions of PP, neurons that responded most strongly to faces showed the greatest difference in responses between expected and unexpected face trials, a characteristic of error units. Fiser *et al.*[47] observed a similar result using calcium imaging to record neural activity in layer 2/3 of mouse V1 while the mice explored a virtual tunnel with oriented gratings (A or B) spaced out along the walls. They identified a group of V1 neurons, resembling expectation units, that exhibited orientation-selective activity in anticipation of an upcoming grating over the course of exposure. Another group of neurons, resembling error units, exhibited stimulus-evoked activity that was selective for a particular orientation and this selectivity was not substantially altered by experience. When an unexpected grating was shown in the final location, expectation neurons fired as if the predicted grating had been presented, while the activity of the orientation-selective error neurons was driven by the actual stimulus. Strong feature-selective predictive activity was associated with greater suppression of activity in error neurons with corresponding stimulus preferences in trials

where the stimulus matched expectations. Conversely, when expectations were violated, stronger predictive activity preceded enhanced error unit responses. In line with the precision-weighting mechanisms described by PP, Fiser *et al.*[47] reported that the strength of predictive activity in V1 neurons and ACC axons projecting to V1 was greater as the mouse approached a position where the identity of the grating stimulus was reliably stable compared with a position where its identity varied. Finally, when the stimulus in the final position was omitted, a strong V1 response was observed and the response strength positively correlated with the strength of the preceding predictive activity. In fact, several studies of locomotor feedback in rodents have identified subpopulations of V1 neurons that signal the magnitude of the mismatch between expected patterns of visual flow and actual sensory input.[48,123,124] Such activity profiles are consistent with PP and difficult to explain in terms of pure feedforward drive.

Laminar segregation

The laminar segregation of bottom-up and top-down activity is an established feature of cortical anatomy.[164] Forward signals are primarily transmitted from superficial to middle layers, while backward connections terminate in superficial and deep layers (Fig. 3A).[166–168] The recent advent of high-resolution neuroimaging has made it possible to exploit this laminar architecture to dissociate putative prediction and error signals in human subjects. Using 7T fMRI, Kok *et al.*[169] found selective increases in activation of the deep layers of V1 induced by illusory figures associated with a Kanisza triangle, while dampened activity was observed in regions of superficial and middle layers of V1 associated with the encoding of the Pac-Man inducers that resembled the laminar profile of genuine bottom-up stimulation. This appears to accord well with PP: expectations are relayed from higher-level areas to deep layers of V1 and the

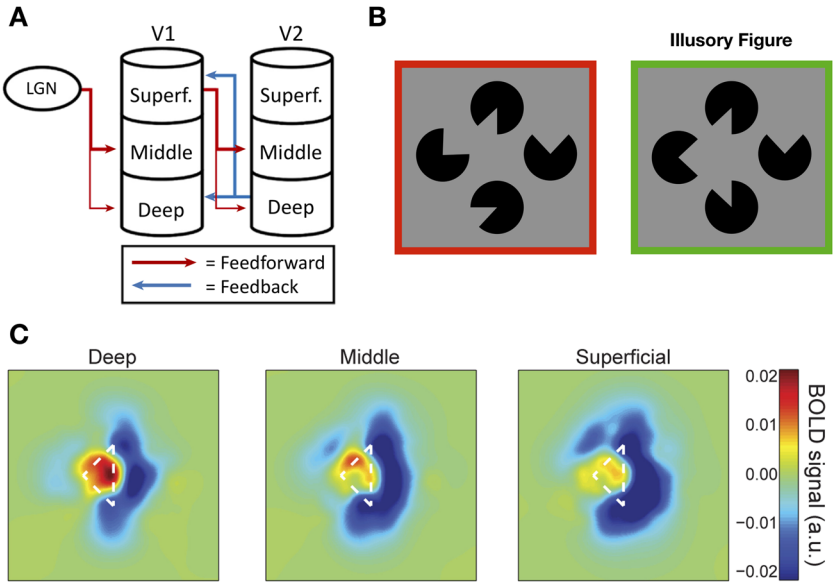


Figure 3. (A) The laminar segregation of feedforward and feedback connections between human lateral geniculate nucleus, V1, and V2. According to PP, expectation units are expected to primarily occupy deep cortical layers, while error units should be found in superficial layers (adapted from Ref. 214). (B) Kok *et al.*[169] presented participants with a Kanizsa illusion while recording their BOLD response with 7T fMRI. The task is well poised to isolate prediction signals because the bottom-up sensory input is the same whether the illusory triangle is presented or not. (C) Data from Kok *et al.* demonstrating the pattern of BOLD activity across different cortical layers in response to a Kanizsa triangle illusion and its "Pac-Man" inducers. The results showed significantly enhanced activity in deep layers of V1 corresponding to the illusory triangle, which could be interpreted as representing the activity of expectation units signaling the presence of the illusory triangle. There was also evidence of suppressed activity in middle and superficial layers of V1 where the Pac-Man inducers fell within the receptive fields. This response suppression exhibited the same laminar profile as the response to a checkerboard stimulus (adapted from Ref. 169).

error units' response to the predictable inducers is suppressed in the same layers as the response to the actual visual stimuli (Fig. 3C). However, it is not clear why the absence of bottom-up sensory input at the location of the illusory contours did not itself evoke a prediction error response in the superficial and middle layers. In another high-resolution fMRI study, Muckli *et al.*[132] showed that only the superficial layers of V1 were significantly activated by a subregion of a visual scene that had been occluded (Fig. 6B). The fact that no subjects exhibited significant activity in the midlayers, where forward information peaked, suggests that this activity may represent descending predictions about the omitted scene. Interestingly, however, there was little evidence of a representation of the scene in deep layers where expectation units are thought to predominate. Although these studies demonstrate the promise of laminar fMRI as a novel means of investigating PP, neither can be said to offer concrete (dis)confirmatory evidence.

Oscillations

In the primate visual system, signals transmitted forward through hierarchical levels are expressed in theta-band (4–8 Hz) and gamma-band (>30 Hz) activity, while backward signaling is carried out through beta-band activity (12–30 Hz).[42,170] PP associates this oscillatory segregation between forward and backward processing[171] with the distinct subpopulations housing error and expectation units, suggesting that oscillatory signatures of error and prediction message passing should be dissociable.[40,42,172] Consistent with the proposed laminar profile of PP dynamics, evidence from monkey and human studies indicate that gamma-band activity predominantly emanates from forward projections in superficial layers, while descending activity, carried by alpha/beta-band activity, is strongest in layers 5/6.[40,170,173–175] Studies that have sought to test PP's characterization of oscillatory message passing have found that when a stimulus is expected, beta power gradually builds in the lead up to stimulus onset, and gamma activity is reduced when those expectations are realized.[45,176,177] Conversely, the violation of expectations is typically associated with increases in gamma power, while beta oscillations are initially reduced before resynchronizing.[172] For example, Arnal *et al.*[84] found that audio–visual mismatches evoked increased gamma activity in auditory cortex, which scaled with the strength of the predictive information. Fujioka *et al.*[178] reported that beta-band activity diminished after each beat of a rhythmic tone and reached a peak just before the next beat, suggesting it may be entrained to the tempo of the beat. When the tone was omitted, the beta-band activity did not decrease, perhaps reflecting the recalibration of temporal inferences, and a sudden peak in gamma-band activity was observed.

A number of other studies have provided evidence consistent with the idea that gamma- and beta-band activity constitute prediction error and expectation signals, respectively. For instance, analyses of Granger causality by Richter *et al.*[179] indicated that top-down beta-band activity targeting macaque V1 enhanced subsequent, spatially specific, stimulus-driven ascending gamma-band signals from V1 to V4 in a manner consistent with precision-weighting mechanisms invoked by PP. Similarly, Sedley *et al.*[180] recorded local field potentials using electrocorticography (ECoG) in three human subjects listening to auditory sequences with occasionally deviant fundamental frequencies. They demonstrated that human gamma-band activity showed a positive correlation with an estimate of precision-weighted prediction error derived from a Bayes-optimal model, while beta-band activity correlated positively with a model estimate of prediction updating. A recent study by Chao *et al.*[158] isolated three components in macaque ECoG recordings of novelty responses in an auditory paradigm, whereby both established local and global regularities can be respected or violated (Fig. 4). They identified two gamma-band components that closely matched model estimates for lower- and higher-level prediction error signals, as well as a beta-band component interpreted as a prediction update signal. When both local and global regularities were violated, the strength of the prediction update signal significantly affected the lower- and higher-level prediction error signals on the subsequent trial. When just the global regularity was violated, prediction update signals only influenced higher-level prediction error signals on the next trial. Chao *et al.* interpreted these findings as evidence that oscillations of distinct frequency channels transmit signals from expectation and error units across the cortical hierarchy. However, there are also some notable discrepancies between these studies. For instance,

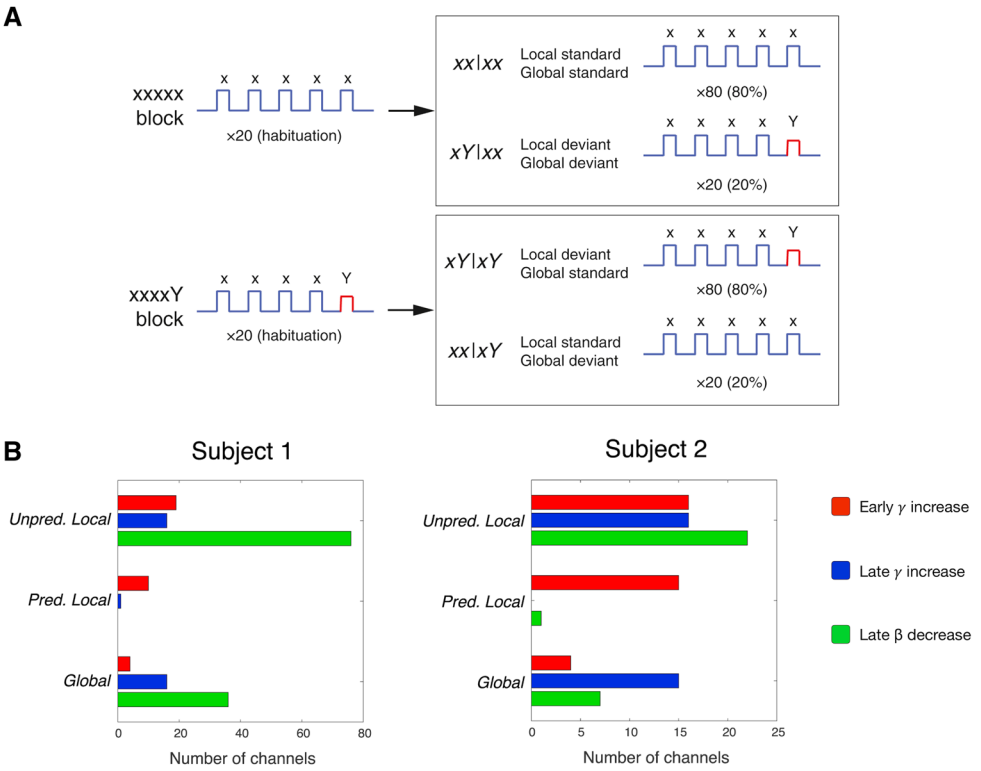


Figure 4. **Experimental paradigm and resulting data from Chao *et al.* (adapted from Ref. 158). (A) Chao *et al.* used an auditory local–global paradigm, whereby local and global regularities could be respected or violated, to investigate the dynamics of oscillatory activity in macaque monkeys using ECoG recordings. Let us assume the standard (expected) sequence is xxxxY. If the sequence presented to the subject is xxxxY, "Y" represents a local deviant (differing from the preceding tone, but not the global pattern). If the sequence played was xxxxx instead, the final "x" represents a global deviant (differing from the global pattern, but not the local pattern). (B) The results showed three components that closely matched model estimates for a low-level prediction error, a higher-level prediction error, and a prediction update signal. The low-level prediction error was associated with an early gamma-power increase over primary auditory cortex and was elicited by unpredicted local deviants. The higher-level prediction error was composed of a late phase gamma-power increase over anterior temporal cortex and was elicited by local and global deviants. Finally, the prediction update component was represented by a late beta-power decrease over prefrontal cortex, which followed the higher-level prediction error, and was associated with unpredicted local and global deviants. The authors found that when both local and global regularities were violated, the activation of the prediction update signal significantly affected activation levels of the low-level prediction error and higher-level prediction error on the subsequent trial. When only the global regularity was violated, the prediction update component's activation only influenced the higher-level prediction error on the next trial.**

Sedley *et al.* reported a positive relationship between beta power and prediction updates, whereas Chao *et al.* found that prediction updates were associated with a reduction in beta power. Additionally, Sedley *et al.* associated prediction precision with alpha-band activity, while Richter *et al.* linked enhanced precision to elevated beta-band influences.

Other auditory studies have not observed these distinct oscillatory signatures or have reported conflicting findings. Using ECoG recordings, El Karoui *et al.*[181] found that local deviants (i.e., a tone differing from the preceding tone in a sequence; see Fig. 4A) evoked high-gamma responses in temporal cortex, consistent with previous studies, but global deviants elicited sustained decreases in beta activity. However, as the task in this study was to identify global deviants, it is difficult to ascertain how much of this effect is attributable to prediction and how much to changes in attention. Todorovic *et al.*[45] found greater gamma power for unexpected tone repetitions and unexpected omissions, but unexpected repetitions also elicited greater power in low frequencies (5–9 Hz). In a

subsequent study, Todorovic *et al.*[106] found that beta activity decreased after unexpected tones, like Chao *et al.*, but only in the absence of attention. If gamma-band oscillations represent the output of error units, unpredictable deviants would be expected to provoke strong gamma-band activity, while predictable deviants should elicit a more muted response. Although Durschmid *et al.*[182] did find that high-gamma (>60 Hz) responses differentiated between unpredictable and predictable deviant tones at frontal sites, unpredictable and predictable deviants elicited equivalent high-gamma activity over temporal cortex.

Hypothesis 3: Summary and conclusions

PP's specification of distinct neural subpopulations firmly divorces the theory from traditional models of sensory processing but also raises a critical methodological hurdle for those seeking to test its tenets. As highlighted above (Hypothesis 1), the question of distinct neural subpopulations presents a serious complication for the interpretation of global, noninvasive brain recording data and, therefore, in this section we have considered single-unit recording studies that have attempted to dissociate these subpopulations. In fact, this issue also presents significant challenges for direct recording studies: if prediction and error units do exist, then definitively testing PP hinges on first ascertaining their prevalence in the sampled population. Acquiring recordings from distinct cortical layers, whether directly or indirectly, may be essential to overcoming many of these obstacles and adding to what is currently a limited evidence base supporting the existence of functionally distinct subpopulations (e.g., Refs. 47 and 98).

Although the assignment of forward and backward pathways to superficial and deep layers respectively is likely an oversimplification,[172,183] there is evidence that where these streams lie in adjacent layers, or even within the same layers, they remain segregated.[164] There is also evidence of a segregation of prediction and error processing into oscillatory bands, but this finding is not universal and some of the discrepancies in the literature are difficult to reconcile.[42] Moreover, it is not known precisely what kind of information is carried in these channels and there is little work describing their mutual influences.[158] Furthermore, while gamma- and beta-band activity have been proposed to support PP mechanisms, other frequency channels may also contribute to this activity.[172,180] Indeed, if PP is a canonical computation, *all* frequency channels would represent some form of inferential processing. Finally, it should be noted that while the presence of subpopulations of error and expectation units is a necessary characteristic of PP, the laminar and oscillatory asymmetries are also compatible with alternative functional architectures (e.g., Ref. 168) and even vary across some implementations of PP (see Ref. 103).

Hypothesis 4: Prediction error minimization is achieved by an inferential hierarchy

As the sections above outline, many studies have identified expectation-related modulations of neural activity that are consistent with PP. However, a central contention of PP models is that perceptual processing is coordinated by an inferential hierarchy.[17] This hierarchy allows regularities at different spatiotemporal scales to exert an influence at an appropriate level of processing, funneling processing toward an internally consistent, globally coherent representation with each iteration. Top-down influences are increasingly recognized as an integral component in early sensory processing,[41,128,184–187] but the contention that these hierarchical dynamics represent prediction error minimization has rarely been explicitly tested.

One avenue for investigating such a framework is to consider the relative latencies of expectation effects. PP suggests that RS occurs because a repeated stimulus becomes expected and, therefore, RS and ES are framed as arising from a common mechanism but one that is applied at distinct levels of the processing hierarchy.[27] Although ES is characterized as a higher-level effect, sensitive to complex regularities over extended periods of integration, RS is considered an automatic expression of low-level prediction error minimization, sensitive to simple, transitional statistics.[55,188] Consistent with PP, numerous studies using MEG/EEG, fMRI, and ECoG show that while local deviants and simple regularities (e.g., repetitions) modulate early phases of neural responses, evidence of sensitivity to more complex sensory inference (e.g., global deviants and ES) emerges at later latencies.[46,69,81,158,181,189–193]

Accordingly, early responses to local deviants are typically confined to primary sensory cortex, while violations of complex regularities evoke activity across distributed, higher-level areas (e.g., Ref. 182; see Ref. 42 for a review). However, it is not clear to what extent these effects can be interpreted as confirmation of a hierarchical PP network.[42] For example, there is substantial uncertainty about what exactly RS represents[27,72,194] and PP and adaptation offer apparently indistinguishable accounts of expectation-independent RS effects. Therefore, the observation that repetition- and expectation-related suppression arise from lower and higher levels of the cortical hierarchy cannot be taken as definitive support for PP.

Although there is evidence that the predictability of a stimulus modulates activity in low-level sensory areas[103,128] and alters the functional connectivity between neural regions associated with its processing,[75,83,155,195] efforts to trace the diffusion of these effects across the processing hierarchy have produced mixed results.[60–62,67,79] For example, Utzerath *et al.*[67] found evidence of RS, ES, and an interaction between expectation and repetition in area LOC. No evidence of RS, nor an interaction between repetition and expectation, was observed in V1, while ES was only detected in right V1. However, it is not clear that predictions about the kinds of complex object images used in these studies should be expected to suppress activity at the earliest stages of sensory processing (Fig. 6D). There is also a reasonable concern that stimuli labeled "unexpected" are often better described as "less expected." In the Utzerath *et al.* paradigm, the stimulus pool was composed of just four stimuli and the "unexpected" transitions occurred in 25% of trials. Given the limited range of outcomes, it may be that the stimulus probabilities had a marginal effect on processing and the associated early visual responses.

Another intriguing question raised by PP is how does a higher-level area, with its own stimulus preferences, communicate predictions about activity at a preceding level of the hierarchy with distinct response properties? Recent invasive neural recordings have attempted to capture the unfolding of these hierarchical interactions over time. Schwiedrzik and Freiwald[92] sought to identify signatures of this process across three levels of the macaque face-patch system, having trained monkeys to associate nine pairs of face stimuli with differing facial identities and head orientations. These three areas are distinguished by their increasingly abstract stimulus preferences: ML, whose neurons exhibit viewpoint specificity; AL, which is characterized by mirror-symmetric response preferences (i.e., profile-selective); and AM, which is tuned to view-independent stimulus identity.[73] Since minimizing the error response to a view violation may require only the revision of sensory expectations at the lowest level of the hierarchy, these errors might be expected to be resolved more quickly than those requiring updates to higher-level predictions. Indeed, the results showed that ML responses to simple violations of head orientation diminished more quickly than more complex stimulus violations involving discrepant facial identities (Fig. 5). In addition, the authors found that mirror-symmetric view violations produced smaller responses in ML than nonmirror-symmetric violations in the late phase of the response. Together, these results could be interpreted as evidence that early responses reflect the local tuning properties of error units, while activity in the later phase of a response represents the involvement of higher-level areas (AL) in the resolution of more complex prediction errors. However, an important detail of this study is that the deviant stimulus pairings were not coded according to the departure of the second face from the expected second face but as the departure of the first face from the normal predictor of the second face. Because the deviant predictor stimuli would, on a subset of trials, also predict a different view to that which appeared, many of the trials coded as "identity violations," could be classed as "identity and view violations," which may offer an alternative explanation for the similarity between the average response to these trial types. Moreover, some of the trials in the "view violation" condition required predictors that were not among the pool of trained predictor stimuli and may therefore have failed to evoke any specific expectations.

In another study, Issa *et al.*[196] pointed out that PP also diverges from other accounts in predicting enhanced activity over time from low-level error units that prefer the presented stimulus features, when the stimulus conflicts with sensory predictions from higher levels. To test this, they simultaneously recorded neuronal activity from

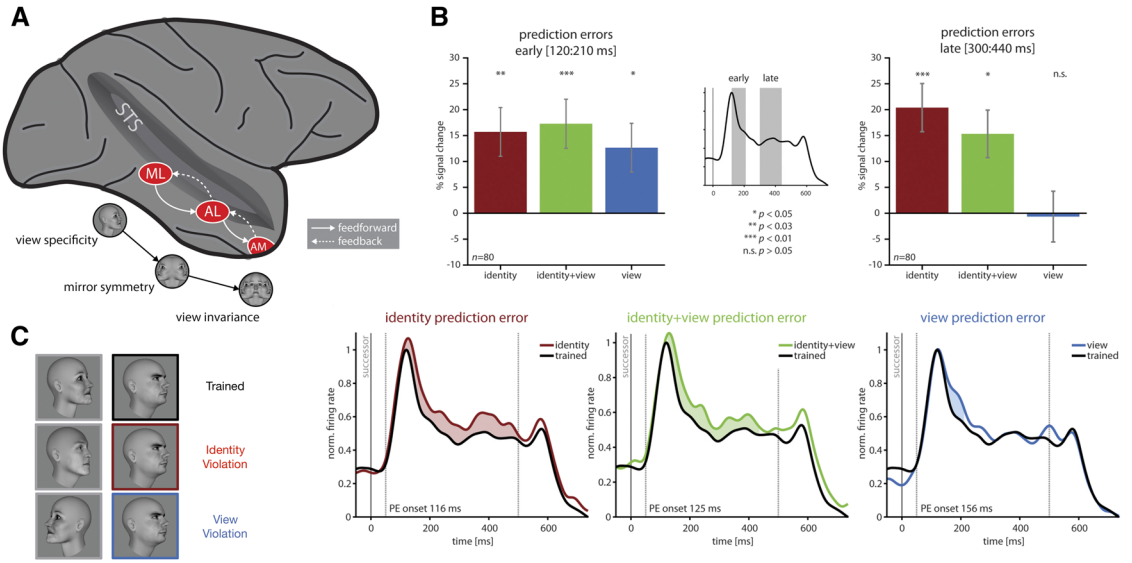


Figure 5. **The macaque face processing system in IT and prediction error responses elicited from its constituent cortical areas (this figure was adapted from Ref. 92). (A) The macaque face processing network is composed of three face-selective areas, each with its own functional specialization, arranged in a three-level processing hierarchy. Although tuning to head orientation decreases from area ML via area AL to area AM, selectivity for facial identity across head orientations increases.[73] (B) Schwiedrzik and Freiwald[92] recorded single-unit activity in ML in response to learned face pairs, where occasional deviant stimuli had a different identity, head position, or both. Although all three violations individually produced a significant early prediction error response, the late sustained response to orientation violations dwindled and disappeared, but the response to identity and combined violations remained significant. These findings suggest that early responses reflect mismatches along local tuning properties, but later responses reflect view-invariant identity errors. Mirror-symmetric view violations evoked smaller prediction errors in ML than nonmirror-symmetric violations in the late phase. As the timing of this effect overlaps with the peak of mirror-symmetric identity tuning in AL, this suggests that higher order representations are involved in suppressing prediction errors in the late phase. (C) However, the interpretation of these results is complicated by the coding scheme where the deviant stimuli were successors to untrained predictors rather than unexpected successors to trained predictors.**

three hierarchical levels of the macaque face processing system while presenting typical and atypical configurations of macaque facial features. The results showed that neurons in lower-level face-selective subregions of IT increasingly responded to atypical configurations of facial features in the late phase of the response, while the higher-level area maintained a preference for typical facial configurations throughout. Furthermore, the late-phase response in the lowest level was better predicted by the early response in higher levels than the lowest level itself, consistent with the idea that updated sensory expectations are distributed from higher levels to resolve lower-level error responses. Issa *et al.* went on to show that these hierarchical neural dynamics could not be captured by feedforward models incorporating lateral inhibition, normalization, or additional excitatory feedback representing Bayesian inference, but were consistent with a PP model describing the lower-level late-phase activity as a prediction error response.

According to PP, the conditional probabilities afforded by higher-level representations may be exploited to constrain lower-level sensory representations, promoting contextual congruence across levels of the hierarchy.[4,132] Indeed, there is evidence that the neural representation of simple stimulus features is influenced by higher-level object representations[115,129] and object representations are modulated by scene context.[197] For example, Harrison *et al.*[198] found that the global coherent motion of a random-dot kinematogram produced diminished responses in V1. Importantly, the constituent dots were deliberately spaced to extend beyond the classical receptive field of a V1 neuron, thus excluding lateral interactions as a potential explanation of this contextual suppression. Similar to the paradigm used by Issa *et al.*,

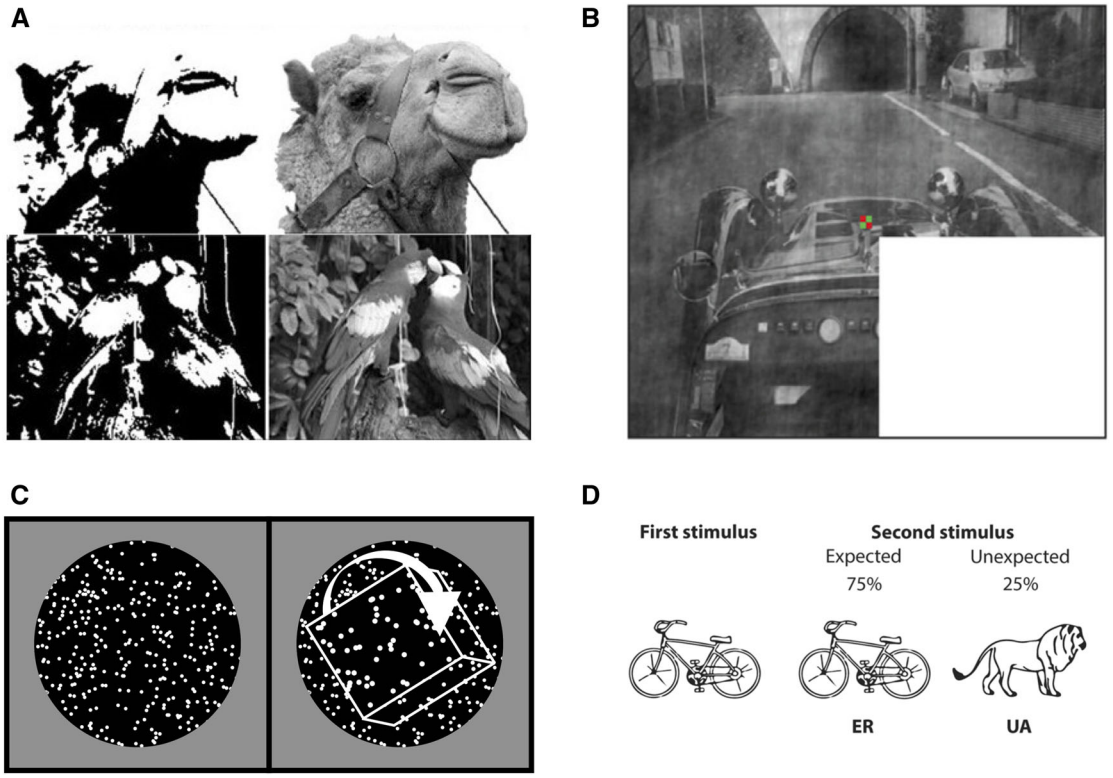


Figure 6. **Examples of stimuli used to probe the nature of sensory prediction. (A) Mooney images used by Hsieh *et al.* (adapted from Ref. 199). The two tone image becomes interpretable when the full greyscale image is seen. (B) An occluded visual scene used by Muckli *et al.* (adapted from Ref. 132). Their analyses were conducted on V1 neurons whose receptive field fell within the occluded region. (C) An illustration of the structure-from-motion stimulus used by Murray *et al.*[115] (D) Expected and unexpected stimulus pairs used by Utzerath *et al.* (adapted from Ref. 67).**

Hsieh *et al.*[199] compared BOLD responses in early visual cortex with Mooney images (Fig. 6A) before and after human subjects recognized the stimulus as a meaningful image. They reported that when subjects could interpret what the two-tone image represented, the response to this stimulus more closely resembled the response to a full grayscale version of the same image than the response to the identical two-tone image before it was interpretable.

Similarly, others have found that the grouping of moving stimulus elements into a perceptual object is associated with increased activity in area LOC and suppressed activity in lower-level areas hMT and V1 (Fig. 6C).[114,115] This observation accords with a predictive model in LOC transmitting its perceptual hypothesis to lower-level areas to quell prediction error. Interestingly, these studies demonstrate prediction effects that complement those observed by Utzerath *et al.* in the same brain regions (see above). One explanation for a more modest effect of ES in V1 observed by Utzerath *et al.* compared with the robust effects reported by Murray *et al.*[115] and Fang and He[114] is that the conditions in these studies differed in terms of the presence or absence of a high-level perceptual hypothesis rather than in terms of a more differential effect of expectation (75% likely versus 25% likely). Disrupting activity in V1/V2 and area LOC using transcranial magnetic stimulation (TMS), Wokke *et al.*[187] found that the critical window for disrupting a Kanizsa illusion was earlier in LOC than V1/V2, suggesting early influences from backward connections play an essential role. Similarly, V1 activity representing an apparent motion illusion (e.g., Ref. 125) is influenced by early feedback from MT,[200] and the detection advantage associated with targets predicted by the apparent motion trace is eliminated by applying TMS to MT before target onset.[186]

Hypothesis 4: Summary and conclusions

Many studies that identify expectational modulations of neural activity struggle to demonstrate systematic effects across contiguous levels of the hierarchy (e.g., Ref. 78). However, as mentioned above, this may be partly due to the fact that it is highly challenging to establish precisely what, if any, predictions may be applied at a given processing level. For example, complex stimuli like faces may not yield any specific predictions at the level of detail pertaining to the simple stimulus preferences of V1 receptive fields.[67,79] Indeed, most of the studies that report PP-like dynamics in V1 use simple stimuli like gratings or flashes.[51,78,107,135,201] PP's stance on this problem is somewhat unclear. PP does not predict that the evidence for perceptual inference should be modular, siloed in an area of the brain specialized for the particular stimulus in question. However, more work is needed to understand how sensory predictions might be conveyed across processing levels[41] and to demarcate the extent to which predictive activity should be detectable as one moves away from those specialized processing hubs.[202]

Overall, there is good evidence that the sensory brain is hierarchically organized and that it exploits descending predictive activity to render sensory signals into meaningful constructs. However, the evidence for a PP inferential hierarchy specifically is scarcer. Although this may reflect on the veracity of the theory, it may also be attributable to the great difficulty such a task poses.[161] It is not sufficient to demonstrate that descending connections to early areas introduce contextual modulations that are essential for perception because this is accommodated by traditional models of visual processing (e.g., Ref. 131). What is required are further demonstrations of PP dynamics unfolding across simultaneous recordings from multiple cortical areas, such that the hierarchical interplay of prediction error and prediction can be verified (e.g., Ref. 196). Efforts to increasingly approximate this standard of evidence may also shed light on the important question of how different neural areas communicate prediction in meaningful ways.[203]

Discussion

Traditional theories of perceptual processing have failed to adequately explain why sensory cortices are infused with masses of descending connections.[10,32,204] For example, backward projections from V2 to V1 are 10 times more numerous than forward projections from the lateral geniculate nucleus to V1[205] and the processing of ascending sensory signals is estimated to account for only 1–2% of the brain's energy consumption.[117] Conversely, PP provides a satisfying explanation of these copious backward connections, impugning their consignment to mere recurrent modulations in feedforward schema.[183,206] The explanatory power of this framework, exemplified in simulations of PP that readily account for phenomena ranging from V1 neuron response properties[207] to bistable perception[29] to perceptual illusions,[30,31] has led some to argue that everything the brain does can ultimately be explained in terms of prediction error minimization.[9,17,32] Such a bold contention has inevitably drawn skepticism, much of which has centered on the theory's empirical foundation.

Without question, PP can comfortably accommodate many neurophysiological observations in research examining the role of expectation in perception. But, in reality, few studies have set out to explicitly test hypotheses that are unique to PP and many empirical observations interpreted as supporting PP are derived from paradigms ill-equipped to discriminate among competing models. A key goal of this review was to highlight a number of recent neurophysiological investigations that have leveraged novel paradigms and technological advances in order to put PP to the test. Although on the whole it must be concluded that the empirical support for each of PP's key hypotheses is mixed, there is much work yet to be done and it is striking that clear-cut counterevidence has yet to emerge.

In fact, a common criticism of PP is that it is very difficult to falsify.[34] PP's specification at the level of an algorithm can be seen as a virtue, providing a framework for the integration of explanations at multiple levels from synapse to processing hierarchy,[7,208] but it can also pose a challenge as researchers seeking to test the theory's validity confront the often murky translation from algorithm to biophysical implementation. Indeed, critiques of the standard implementation of PP,[18] focusing on the model's biological plausibility (e.g., neurons with both positive and negative firing rates), have spawned numerous alternative versions with revised circuitries and diverging predictions.[2,41] Confusion can sometimes arise where it is not

clear which of these various incarnations are being tested in a particular study. In addition, the flexibility conferred on PP models by hypothesizing functionally distinct neural subpopulations and layering a precision-weighting mechanism on their responses at every level of the hierarchy renders the theory capable of accommodating apparently contradictory results.[34]

A necessary next step, therefore, is to provide definitive evidence of the existence of expectation and error units in neural processing. Although this represents a considerable challenge using conventional paradigms and neurophysiological techniques, methodological advances in high-resolution fMRI, optogenetics, calcium imaging, and serial single-unit recordings at multiple levels of the processing hierarchy are providing powerful new opportunities to trace neural markers of hierarchical PP dynamics.[41,128,158,196,209,210] Pairing these increasingly sophisticated neural assays with anatomical models, computational modeling, and simulations[137,211] will enable researchers to derive fine-grained *a priori* hypotheses and compare model evidence for variant architectures and also for near-variant ones that share much with the core PP picture but differ in their conceptions of the encoding, flow, or use of prediction errors (e.g., Refs. 2, 10 and 11; see Ref. 137). As the theory is refined by these more delicate neurophysiological tests we hope to see a bridging of the gaps between parallel literatures that currently exist somewhat in isolation[212] (e.g., perceptual decision making).

Although the debate about PP's empirical grounding is currently unsettled, the theory can nevertheless be regarded as a milestone in cognitive neuroscience, spurring efforts to recognize the importance of backward connections in the architecture of the neocortex and the role of prediction in sensory processing.[206,213] Does PP reflect a questionable commitment to bringing multiple phenomena under a single unifying umbrella—one that may not be able to do full justice to any of them? Or can a mature PP capture the full sweep of behavioral effects and experimental data, revealing them as flowing from a core rationale expressed using a handful of repeated processing motifs? Whatever the outcome, this is a lively and ever-expanding literature that allows us to revisit many long-standing assumptions regarding neural function, the nature of mind, and the origins of human behavior.

Acknowledgments

K.S.W. was supported by an Irish Research Council Government of Ireland Postgraduate Scholarship—GOIPG/2017/1093. A.C. was supported by Horizon 2020 European Union ERC Advanced Grant XSPECT—DLV-692739. R.G.O. was supported by Horizon 2020 European Research Council Starting Grant Human Decisions—63829.

Competing interests

The authors declare no competing interests.

References

1. Knill, D.C. & A. Pouget. 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* **27:** 712–719.
2. Spratling, M.W. 2017. A review of predictive coding algorithms. *Brain Cogn.* **112:** 92–97.
3. Summerfield, C. & T. Egner. 2009. Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* **13:** 403–409.
4. de Lange, F.P., M. Heilbron & P. Kok. 2018. How do expectations shape perception? *Trends Cogn. Sci.* **22:** 1–16.
5. Purves, D., Y. Morgenstern & W.T. Wojtach. 2015. Perception and reality: why a wholly empirical paradigm is needed to understand vision. *Front. Sys. Neurosci.* **9:** 18–10.
6. Seriès, P. & A. Seitz. 2013. Learning what to expect (in visual perception). *Front. Hum. Neurosci.* **7:** 668.
7. Lochmann, T. & S. Denève. 2011. Neural processing as causal inference. *Curr. Opin. Neurobiol.* **21:** 774–781.
8. Pouget, A., J.M. Beck, W.J. Ma, *et al*. 2013. Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* **16:** 1170–1178.
9. Friston, K.J. 2010. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11:** 127–138.
10. Heeger, D.J. 2017. Theory of cortical function. *Proc. Natl. Acad. Sci. USA* **114:** 1773–1782.
11. Trapp, S. & M. Bar. 2015. Prediction, context, and competition in visual recognition. *Ann. N.Y. Acad. Sci.* **1339:** 190–198.
12. Hubel, D.H. & T.N. Wiesel. 1968. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **195:** 215–243.
13. Riesenhuber, M. & T. Poggio. 1999. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2:** 1019–1025.
14. DiCarlo, J.J., D. Zoccolan & N.C. Rust. 2012. How does the brain solve visual object recognition? *Neuron* **73:** 415–434.
15. Ramachandran, V.S. 1990. Visual perception in people and machines. In *AI and the Eye*. A. Blake & T. Troscianko, Eds.: 21–77. New York: John Wiley & Sons.
16. Beck, J.M., P.E. Latham & A. Pouget. 2011. Marginalization in neural circuits with divisive normalization. *J. Neurosci.* **31:** 15310–15319.
17. Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36:** 181–204.

18. Rao, R.P.N. & D.H. Ballard. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2:** 79–87.
19. Summerfield, C. & F.P. de Lange. 2014. Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* **15:** 745–756.
20. Wiese, W. & T. Metzinger. 2017. Vanilla PP for philosophers: a primer on predictive processing. In *Philosophy and Predictive Processing*. T. Metzinger & W. Wiese, Eds.: 1–18. Frankfurt am Main: MIND Group.
21. Ouden den, H.E.M., P. Kok & F.P. de Lange. 2012. How prediction errors shape perception, attention, and motivation. *Front. Psychol.* **3:** 548.
22. Parr, T. & K.J. Friston. 2017. The active construction of the visual world. *Neuropsychologia* **104:** 92–101.
23. Sims, A. 2017. The problems with prediction—the dark room problem and the scope dispute. In *Philosophy and Predictive Processing*. T. Metzinger & W. Wiese, Eds.: 1–18. Frankfurt am Main: MIND Group.
24. Brown, H.R. & K.J. Friston. 2012. Dynamic causal modelling of precision and synaptic gain in visual perception—an EEG study. *Neuroimage* **63:** 223–231.
25. Kanai, R., Y. Komura, S. Shipp, *et al*. 2015. Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**. https://doi.org/10.1098/rstb.2014.0169.
26. Clark, A. 2017. A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenol. Cogn. Sci.* **17:** 521–534.
27. Auksztulewicz, R. & K.J. Friston. 2016. Repetition suppression and its contextual determinants in predictive coding. *Cortex* **80:** 125–140.
28. Feldman, H. & K.J. Friston. 2010. Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* **4:** 215.
29. Weilnhammer, V., H. Stuke, G. Hesselmann, *et al*. 2017. A predictive coding account of bistable perception—a model-based fMRI study. *PLoS Comput. Biol.* **13**. https://doi.org/10.1371/journal.pcbi.1005536.
30. Lotter, W., G. Kreiman & D. Cox. 2018. A neural network trained to predict future video frames mimics critical properties of biological neuronal responses and perception. arxiv.org.
31. Watanabe, E., A. Kitaoka, K. Sakamoto, *et al*. 2018. Illusory motion reproduced by deep neural networks trained for prediction. *Front. Psychol.* **9:** 1143–1112.
32. Clark, A. 2019. Beyond desire? Agency, choice, and the predictive mind. *Australas. J. Philos.* **9:** 1–15.
33. Egner, T. & C. Summerfield. 2013. Grounding predictive coding models in empirical neuroscience research. *Behav. Brain Sci.* **36:** 210–211.
34. Kogo, N. & C. Trengove. 2015. Is predictive coding theory articulated enough to be testable? *Front. Comput. Neurosci.* **9:** 111.
35. Ransom, M., S. Fazelpour & C. Mole. 2017. Attention in the predictive mind. *Conscious. Cogn.* **47:** 99–112.
36. Block, N. 2018. If perception is probabilistic, why does it not seem probabilistic? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373**. https://doi.org/10.1098/rstb.2017.0341.
37. Wacongne, C., J.P. Changeux & S. Dehaene. 2012. A neuronal model of predictive coding accounting for the mismatch negativity. *J. Neurosci.* **32:** 3665–3678.
38. Lee, T.S. & D. Mumford. 2003. Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.* **20:** 1434–1448.
39. Friston, K.J. 2005. A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360:** 815–836.
40. Bastos, A.M., W.M. Usrey, R.A. Adams, *et al*. 2012. Canonical microcircuits for predictive coding. *Neuron* **76:** 695–711.
41. Keller, G.B. & T.D. Mrsic-Flogel. 2018. Predictive processing: a canonical cortical computation. *Neuron* **100:** 424–435.
42. Heilbron, M. & M. Chait. 2018. Great expectations: is there evidence for predictive coding in auditory cortex? *Neuroscience* **389:** 54–73.
43. Bendixen, A., E. Schröger & I. Winkler. 2009. I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. *J. Neurosci.* **29:** 8447–8451.
44. SanMiguel, I., K. Saupe & E. Schröger. 2013. I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted "what" but not "when." *Front. Hum. Neurosci.* **7:** 407.
45. Todorovic, A., F. van Ede, E. Maris, *et al*. 2011. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J. Neurosci.* **31:** 9118–9123.
46. Wacongne, C., E. Labyt, V. van Wassenhove, *et al*. 2011. Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci. USA* **108:** 20754–20759.
47. Fiser, A., D. Mahringer, H.K. Oyibo, *et al*. 2016. Experience-dependent spatial expectations in mouse visual cortex. *Nat. Neurosci.* **19:** 1658–1664.
48. Keller, G.B., T. Bonhoeffer & M. Hübener. 2012. Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* **74:** 809–815.
49. Kok, P., M.F. Failing & F.P. de Lange. 2014. Prior expectations evoke stimulus templates in the primary visual cortex. *J. Cogn. Neurosci.* **26:** 1546–1554.
50. May, P.J.C. & H. Tiitinen. 2010. Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology* **47:** 66–122.
51. Ekman, M., P. Kok & F.P. de Lange. 2017. Time-compressed preplay of anticipated events in human primary visual cortex. *Nat. Commun.* **8:** 15276.
52. Gavornik, J.P. & M.F. Bear. 2014. Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nat. Neurosci.* **17:** 732–737.
53. Kok, P., P. Mostert & F.P. de Lange. 2017. Prior expectations induce prestimulus sensory templates. *Proc. Natl. Acad. Sci. USA* **114:** 10473–10478.
54. Grill-Spector, K., R. Henson & A. Martin. 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* **10:** 14–23.
55. Grotheer, M. & G. Kovács. 2016. Can predictive coding explain repetition suppression? *Cortex* **80:** 113–124.

56. Vogels, R. 2016. Sources of adaptation of inferior temporal cortical responses. *Cortex* **80:** 185–195.
57. Summerfield, C., E.H. Trittschuh, J.M. Monti, *et al*. 2008. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* **11:** 1004–1006.
58. Kovács, G., L. Iffland, Z. Vidnyánszky, *et al*. 2012. Stimulus repetition probability effects on repetition suppression are position invariant for faces. *Neuroimage* **60:** 2128–2135.
59. Kovács, G., D. Kaiser, D.A. Kaliukhovich, *et al*. 2013. Repetition probability does not affect fMRI repetition suppression for objects. *J. Neurosci.* **33:** 9805–9812.
60. Larsson, J. & A.T. Smith. 2012. FMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cereb. Cortex* **22:** 567–576.
61. Grotheer, M. & G. Kovács. 2015. The relationship between stimulus repetitions and fulfilled expectations. *Neuropsychologia* **67:** 175–182.
62. Grotheer, M. & G. Kovács. 2014. Repetition probability effects depend on prior experiences. *J. Neurosci.* **34:** 6640–6646.
63. Mayrhauser, L., J. Bergmann, J. Crone, *et al*. 2014. Neural repetition suppression: evidence for perceptual expectation in object-selective regions. *Front. Hum. Neurosci.* **8:** 277.
64. Costa-Faidella, J., T. Baldeweg, S. Grimm, *et al*. 2011. Interactions between "what" and "when" in the auditory system: temporal predictability enhances repetition suppression. *J. Neurosci.* **31:** 18590–18597.
65. Stefanics, G., M. Kimura & I. Czigler. 2011. Visual mismatch negativity reveals automatic detection of sequential regularity violation. *Front. Hum. Neurosci.* **5:** 46.
66. Summerfield, C., V. Wyart, V.M. Johnen, *et al*. 2011. Human scalp electroencephalography reveals that repetition suppression varies with expectation. *Front. Hum. Neurosci.* **5:** 67.
67. Utzerath, C., E. St John-Saaltink, J. Buitelaar, *et al*. 2017. Repetition suppression to objects is modulated by stimulus-specific expectations. *Sci. Rep.* **7:** 8781.
68. Olkkonen, M., G.K. Aguirre & R.A. Epstein. 2017. Expectation modulates repetition priming under high stimulus variability. *Journal of Vision* **17:** 10.
69. Tang, M.F., C.A. Smout, E. Arabzadeh, *et al*. 2018. Prediction error and repetition suppression have distinct effects on neural representations of visual information. *eLife* **7:** 2108.
70. Pajani, A., S. Kouider, P. Roux, *et al*. 2017. Unsuppressible repetition suppression and exemplar-specific expectation suppression in the fusiform face area. *Sci. Rep.* **7:** 833.
71. Kaliukhovich, D.A. & R. Vogels. 2011. Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cereb. Cortex* **21:** 1547–1558.
72. Vinken, K., H.P. Op de Beeck & R. Vogels. 2018. Face repetition probability does not affect repetition suppression in Macaque inferotemporal cortex. *J. Neurosci.* **38:** 7492–7504.
73. Freiwald, W.A. & D.Y. Tsao. 2010. Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* **330:** 845–851.
74. Ouden den, H.E.M., K.J. Friston, N.D. Daw, *et al*. 2009. A dual role for prediction error in associative learning. *Cereb. Cortex* **19:** 1175–1185.
75. Ouden den, H.E.M., J. Daunizeau, J. Roiser, *et al*. 2010. Striatal prediction error modulates cortical coupling. *J. Neurosci.* **30:** 3210–3219.
76. Egner, T., J.M. Monti & C. Summerfield. 2010. Expectation and surprise determine neural population responses in the ventral visual stream. *J. Neurosci.* **30:** 16601–16608.
77. Garrido, M.I., E.G. Rowe, V. Halász, *et al*. 2018. Bayesian mapping reveals that attention boosts neural responses to predicted and unpredicted stimuli. *Cereb. Cortex* **28:** 1771–1782.
78. Kok, P., J.F.M. Jehee & F.P. de Lange. 2012. Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* **75:** 265–270.
79. Richter, D., M. Ekman & F.P. de Lange. 2018. Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *J. Neurosci.* **38:** 7452–7461.
80. Summerfield, C., T. Egner, M. Greene, *et al*. 2006. Predictive codes for forthcoming perception in the frontal cortex. *Science* **314:** 1311–1314.
81. Todorovic, A. & F.P. de Lange. 2012. Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J. Neurosci.* **32:** 13389–13395.
82. Trapp, S., J. Lepsien, S.A. Kotz, *et al*. 2016. Prior probability modulates anticipatory activity in category-specific areas. *Cogn. Affect. Behav. Neurosci.* **16:** 135–144.
83. Arnal, L.H., B. Morillon, C.A. Kell, *et al*. 2009. Dual neural routing of visual facilitation in speech processing. *J. Neurosci.* **29:** 13445–13453.
84. Arnal, L.H., V. Wyart & A.-L. Giraud. 2011. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat. Neurosci.* **14:** 797–801.
85. Ramachandran, S., T. Meyer & C.R. Olson. 2016. Prediction suppression in monkey inferotemporal cortex depends on the conditional probability between images. *J. Neurophysiol.* **115:** 355–362.
86. Kanwisher, N., J. McDermott & M.M. Chun. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17:** 4302–4311.
87. de Gardelle, V., M. Waszczuk, T. Egner, *et al*. 2013. Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. *Cereb. Cortex* **23:** 2235–2244.
88. Bell, A.H., C. Summerfield, E.L. Morin, *et al*. 2016. Encoding of stimulus probability in macaque inferior temporal cortex. *Curr. Biol.* **26:** 2280–2290.
89. Kaposvari, P., S. Kumar & R. Vogels. 2018. Statistical learning signals in macaque inferior temporal cortex. *Cereb. Cortex* **28:** 250–266.
90. Meyer, T. & C.R. Olson. 2011. Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc. Natl. Acad. Sci. USA* **108:** 19401–19406.
91. Meyer, T., S. Ramachandran & C.R. Olson. 2014. Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *J. Neurosci.* **34:** 9332–9337.

92. Schwiedrzik, C.M. & W.A. Freiwald. 2017. High-level prediction signals in a low-level area of the macaque face-processing hierarchy. *Neuron* **96:** 89–98.
93. Vinken, K. & R. Vogels. 2017. Adaptation can explain evidence for encoding of probabilistic information in macaque inferior temporal cortex. *Curr. Biol.* **27:** R1210–R1212.
94. Bell, A.H., C. Summerfield, E.L. Morin, *et al.* 2017. Reply to Vinken and Vogels. *Curr. Biol.* **27:** R1212–R1213.
95. Anderson, B. & D.L. Sheinberg. 2008. Effects of temporal context and temporal expectancy on neural activity in inferior temporal cortex. *Neuropsychologia* **46:** 947–957.
96. Kaliukhovich, D.A. & R. Vogels. 2014. Neurons in macaque inferior temporal cortex show no surprise response to deviants in visual oddball sequences. *J. Neurosci.* **34:** 12801–12815.
97. Turk-Browne, N.B., B.J. Scholl, M.M. Chun, *et al.* 2009. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J. Cogn. Neurosci.* **21:** 1934–1945.
98. Zmarz, P. & G.B. Keller. 2016. Mismatch receptive fields in mouse visual cortex. *Neuron* **92:** 766–772.
99. Auksztulewicz, R. & K.J. Friston. 2015. Attentional enhancement of auditory mismatch responses: a DCM/MEG study. *Cereb. Cortex* **25:** 4273–4283.
100. Fardo, F., R. Auksztulewicz, M. Allen, *et al.* 2017. Expectation violation and attention to pain jointly modulate neural gain in somatosensory cortex. *Neuroimage* **153:** 109–121.
101. Jiang, J., C. Summerfield & T. Egner. 2013. Attention sharpens the distinction between expected and unexpected percepts in the visual brain. *J. Neurosci.* **33:** 18438–18447.
102. Kok, P., D. Rahnev, J.F.M. Jehee, *et al.* 2012. Attention reverses the effect of prediction in silencing sensory signals. *Cereb. Cortex* **22:** 2197–2206.
103. Kok, P. & F.P. de Lange. 2015. Predictive coding in sensory cortex. In *An Introduction to Model-Based Cognitive Neuroscience*. B. U. Forstmann & E.-J. Wagenmakers, Eds.: 221–244. Berlin: Springer Science + Business Media.
104. Hsu, Y.-F., J.A. Hämäläinen & F. Waszak. 2014. Both attention and prediction are necessary for adaptive neuronal tuning in sensory processing. *Front. Hum. Neurosci.* **8:** 14608.
105. Jaramillo, S. & A.M. Zador. 2011. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nat. Neurosci.* **14:** 246–253.
106. Todorovic, A., J.-M. Schoffelen, F. van Ede, *et al.* 2015. Temporal expectation and attention jointly modulate auditory oscillatory activity in the beta band. *PLoS One* **10:** e0120288.
107. St John-Saaltink, E., C. Utzerath, P. Kok, *et al.* 2015. Expectation suppression in early visual cortex depends on task set. *PLoS One* **10:** e0131172.
108. Hsu, Y.-F., J.A. Hämäläinen & F. Waszak. 2018. The processing of mispredicted and unpredicted sensory inputs interact differently with attention. *Neuropsychologia* **111:** 85–91.
109. Blank, H. & M.H. Davis. 2016. Prediction errors but not sharpened signals simulate multivoxel fmri patterns during speech perception. *PLoS Biol.* **14:** e1002577.
110. Kaliukhovich, D.A., W. De Baene & R. Vogels. 2013. Effect of adaptation on object representation accuracy in macaque inferior temporal cortex. *J. Cogn. Neurosci.* **25:** 777–789.
111. Kumar, S., P. Kaposvari & R. Vogels. 2017. Encoding of predictable and unpredictable stimuli by inferior temporal cortical neurons. *J. Cogn. Neurosci.* **29:** 1445–1454.
112. Yon, D., S.J. Gilbert, F.P. de Lange, *et al.* 2018. Action sharpens sensory representations of expected outcomes. *Nat. Commun.* **9:** 849.
113. Marques, T., J. Nguyen, G. Fioreze, *et al.* 2018. The functional organization of cortical feedback inputs to primary visual cortex. *Nat. Neurosci.* **21:** 757–764.
114. Fang, F., D. Kersten & S.O. Murray. 2008. Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *J. Vis.* **8:** 2–9.
115. Murray, S.O., D. Kersten, B.A. Olshausen, *et al.* 2002. Shape perception reduces activity in human primary visual cortex. *Proc. Natl. Acad. Sci. USA* **99:** 15164–15169.
116. Logothetis, N.K. 2008. What we can do and what we cannot do with fMRI. *Nature* **453:** 869–878.
117. Raichle, M.E. & M.A. Mintun. 2006. Brain work and brain imaging. *Annu. Rev. Neurosci.* **29:** 449–476.
118. Cohen, M.X. 2017. Where does EEG come from and what does it mean? *Trends Neurosci.* **40:** 208–218.
119. Chennu, S., V. Noreika, D. Gueorguiev, *et al.* 2016. Silent expectations: dynamic causal modeling of cortical prediction and attention to sounds that weren't. *J. Neurosci.* **36:** 8305–8316.
120. Schapiro, A.C., L.V. Kustner & N.B. Turk-Browne. 2012. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr. Biol.* **22:** 1622–1627.
121. Dehaene, S., F. Meyniel, C. Wacongne, *et al.* 2015. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* **88:** 2–19.
122. Xu, S., W. Jiang, M.-M. Poo, *et al.* 2012. Activity recall in a visual cortical ensemble. *Nat. Neurosci.* **15:** 449–455.
123. Attinger, A., B. Wang & G.B. Keller. 2017. Visuomotor coupling shapes the functional development of mouse visual cortex. *Cell* **169:** 1291–1302.
124. Leinweber, M., D.R. Ward, J.M. Sobczak, *et al.* 2017. A sensorimotor circuit in mouse cortex for visual flow Predictions. *Neuron* **95:** 1420–1432.e5.
125. Alink, A., C.M. Schwiedrzik, A. Kohler, *et al.* 2010. Stimulus predictability reduces responses in primary visual cortex. *J. Neurosci.* **30:** 2960–2966.
126. Kok, P., G.J. Brouwer, M.A.J. van Gerven, *et al.* 2013. Prior expectations bias sensory representations in visual cortex. *J. Neurosci.* **33:** 16275–16284.
127. Teufel, C. & B. Nanay. 2017. How to (and how not to) think about top-down influences on visual perception. *Conscious. Cogn.* **47:** 17–25.
128. Petro, L.S., L. Vizioli & L. Muckli. 2014. Contributions of cortical feedback to sensory processing in primary visual cortex. *Front. Psychol.* **5:** 1427.

129. Kok, P. & F.P. de Lange. 2014. Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Curr. Biol.* **24:** 1531–1535.
130. Lee, T.S. & M. Nguyen. 2001. Dynamics of subjective contour formation in the early visual cortex. *Proc. Natl. Acad. Sci. USA* **98:** 1907–1911.
131. Zhang, N.R. & R. Von Der Heydt. 2010. Analysis of the context integration mechanisms underlying figure-ground organization in the visual cortex. *J. Neurosci.* **30:** 6482–6496.
132. Muckli, L., F. De Martino, L. Vizioli, *et al.* 2015. Contextual feedback to superficial layers of V1. *Curr. Biol.* **25:** 2690–2695.
133. van Kerkoerle, T., M.W. Self & P.R. Roelfsema. 2017. Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat. Commun.* **8:** 370.
134. Eagleman, S.L. & V. Dragoi. 2012. Image sequence reactivation in awake V4 networks. *Proc. Natl. Acad. Sci. USA* **109:** 19450–19455.
135. Chong, E., A.M. Familiar & W.M. Shim. 2016. Reconstructing representations of dynamic visual objects in early visual cortex. *Proc. Natl. Acad. Sci. USA* **113:** 1453–1458.
136. Saleem, A.B., A.I. Ayaz, K.J. Jeffery, *et al.* 2013. Integration of visual motion and locomotion in mouse visual cortex. *Nat. Neurosci.* **16:** 1864–1869.
137. Spratling, M.W. 2019. Fitting predictive coding to the neurophysiological data. *Brain Res.* **1720:** 146313.
138. Yon, D., F.P. de Lange & C. Press. 2019. The predictive brain as a stubborn scientist. *Trends Cogn. Sci.* **23:** 6–8.
139. Maniscalco, B., J.L. Lee, P. Abry, *et al.* 2018. Neural integration of stimulus history underlies prediction for naturalistically evolving sequences. *J. Neurosci.* **38:** 1541–1557.
140. Girshick, A.R., M.S. Landy & E.P. Simoncelli. 2011. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* **14:** 926–932.
141. Berkes, P., G. Orbán, M. Lengyel, *et al.* 2011. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* **331:** 80–83.
142. Adams, W.J., E.W. Graf & M.O. Ernst. 2004. Experience can change the "light-from-above" prior. *Nat. Neurosci.* **7:** 1057–1058.
143. Sotiropoulos, G., A.R. Seitz & P. Seris. 2011. Changing expectations about speed alters perceived motion direction. *Curr. Biol.* **21:** R883–R884.
144. Barascud, N., M.T. Pearce, T.D. Griffiths, *et al.* 2016. Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proc. Natl. Acad. Sci. USA* **113:** E616–E625.
145. Burr, D.C. & G.M. Cicchini. 2014. Vision: efficient adaptive coding. *Curr. Biol.* **24:** R1096–R1098.
146. Murray, S.O., H. Boyaci & D. Kersten. 2006. The representation of perceived angular size in human primary visual cortex. *Nat. Neurosci.* **9:** 429–434.
147. Sterzer, P., C.D. Frith & P. Petrovic. 2008. Believing is seeing: expectations alter visual awareness. *Curr. Biol.* **18:** R697–8.
148. Li, N. & J.J. DiCarlo. 2008. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science* **321:** 1502–1507.
149. Hsu, Y.-F., F. Waszak & J.A. Hämäläinen. 2019. Prior precision modulates the minimization of auditory prediction error. *Front. Hum. Neurosci.* **13:** 283–289.
150. Todd, J. & R. Cornwell. 2018. The importance of precision to updating models of the sensory environment. *Biol. Psychol.* **139:** 8–16.
151. Garrido, M.I., C.L.J. Teng, J.A. Taylor, *et al.* 2016. Surprise responses in the human brain demonstrate statistical learning under high concurrent cognitive demand. *NPJ Sci. Learn* **1:** 16006. https://doi.org/10.1038/npjscilearn.2016.6. eCollection_2016.
152. Sohoglu, E. & M.C. Elife. 2016. Detecting and representing predictable structure during auditory scene analysis. *eLife* **5:** 1–17.
153. Winkler, I., P. Paavilainen, K. Alho, *et al.* 1990. The effect of small variation of the frequent auditory stimulus on the event-related brain potential to the infrequent stimulus. *Psychophysiology* **27:** 228–235.
154. Southwell, R. & M. Chait. 2018. Enhanced deviant responses in patterned relative to random sound sequences. *Cortex* **109:** 92–103.
155. Auksztulewicz, R., N. Barascud, G. Cooray, *et al.* 2017. The cumulative effects of predictability on synaptic gain in the auditory processing stream. *J. Neurosci.* **37:** 6751–6760.
156. Garrido, M.I. & K.J. Friston. 2009. Dynamic causal modeling of the response to frequency deviants. *J. Neurophysiol.* **101:** 2620–2631.
157. Winkler, I., S.L. Denham & I. Nelken. 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci.* **13:** 532–540.
158. Chao, Z.C., K. Takaura, L. Wang, *et al.* 2018. Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. *Neuron* **100:** 1252–1266.
159. Todd, J., A. Provost, L. Whitson, *et al.* 2018. Initial uncertainty impacts statistical learning in sound sequence processing. *Neuroscience* **389:** 41–53.
160. Bendixen, A., W. Prinz, J. Horváth, *et al.* 2008. Rapid extraction of auditory feature contingencies. *Neuroimage* **41:** 1111–1119.
161. Denham, S.L. & I. Winkler. 2018. Predictive coding in auditory perception: challenges and unresolved questions. *Eur. J. Neurosci.* **66:** 1–10.
162. Auksztulewicz, R., C.M. Schwiedrzik, T. Thesen, *et al.* 2018. Not all predictions are equal: "what" and 'when' predictions modulate activity in auditory cortex through different mechanisms. *J. Neurosci.* **38:** 8680–8693.
163. Rimmele, J., H. Jolsvai & E. Sussman. 2011. Auditory target detection is affected by implicit temporal and spatial expectations. *J. Cogn. Neurosci.* **23:** 1136–1147.
164. Markov, N.T., J. Vezoli, P. Chameau, *et al.* 2013. Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.* **522:** 225–259.
165. Berezovskii, V.K., J.J. Nassi & R.T. Born. 2011. Segregation of feedforward and feedback projections in mouse visual cortex. *J. Comp. Neurol.* **519:** 3672–3683.
166. Felleman, D.J. & D.C. Van Essen. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1:** 1–47.

167. Lawrence, S.J.D., D.G. Norris & F.P. de Lange. 2019. Dissociable laminar profiles of concurrent bottom-up and top-down modulation in the human visual cortex. *eLife* **8:** e44422.
168. Self, M.W., T. van Kerkoerle, H. Supèr, *et al*. 2013. Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Curr. Biol.* **23:** 2121–2129.
169. Kok, P., L.J. Bains, T. van Mourik, *et al*. 2016. Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Curr. Biol.* **26:** 371–376.
170. Bastos, A.M., J. Vezoli, C.A. Bosman, *et al*. 2015. Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* **85:** 390–401.
171. Scheeringa, R. & P. Fries. 2019. Cortical layers, rhythms and BOLD signals. *Neuroimage* **197:** 689–698.
172. Arnal, L.H. & A.-L. Giraud. 2012. Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* **16:** 390–398.
173. Michalareas, G., J. Vezoli, S. van Pelt, *et al*. 2016. Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* **89:** 384–397.
174. van Kerkoerle, T., M.W. Self, B. Dagnino, *et al*. 2014. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl Acad. Sci. USA* **111:** 14332–14341.
175. Xing, D., C.I. Yeh, S. Burns, *et al*. 2012. Laminar analysis of visually evoked activity in the primary visual cortex. *Proc. Natl. Acad. Sci. USA* **109:** 13871–13876.
176. Fujioka, T., B. Ross & L.J. Trainor. 2015. Beta-band oscillations represent auditory beat and its metrical hierarchy in perception and imagery. *J. Neurosci.* **35:** 15187–15198.
177. Saleh, M., J. Reimer, R. Penn, *et al*. 2010. Fast and slow oscillations in human primary motor cortex predict oncoming behaviorally relevant cues. *Neuron* **65:** 461–471.
178. Fujioka, T., L.J. Trainor, E.W. Large, *et al*. 2009. Beta and gamma rhythms in human auditory cortex during musical beat processing. *Ann. N.Y. Acad. Sci.* **1169:** 89–92.
179. Richter, C.G., W.H. Thompson, C.A. Bosman, *et al*. 2017. Top-down beta enhances bottom-up gamma. *J. Neurosci.* **37:** 6698–6711.
180. Sedley, W., P.E. Gander, S. Kumar, *et al*. 2016. Neural signatures of perceptual inference. *eLife* **5:** 797.
181. Karoui El, I., J.-R. King, J. Sitt, *et al*. 2015. Event-related potential, time-frequency, and functional connectivity facets of local and global auditory novelty processing: an intracranial study in humans. *Cereb. Cortex* **25:** 4203–4212.
182. Durschmid, S., E. Edwards, C. Reichert, *et al*. 2016. Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proc. Natl. Acad. Sci. USA* **113:** 6755–6760.
183. Markov, N.T. & H. Kennedy. 2013. The importance of being hierarchical. *Curr. Opin. Neurobiol.* **23:** 187–194.
184. Ruff, C.C., F. Blankenburg, O. Bjoertomt, *et al*. 2006. Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex. *Curr. Biol.* **16:** 1479–1488.
185. Huh, C.Y.L., J.P. Peach, C. Bennett, *et al*. 2018. Feature-specific organization of feedback pathways in mouse visual cortex. *Curr. Biol.* **28:** 114–120.
186. Vetter, P., M.-H. Grosbras & L. Muckli. 2015. TMS over V5 disrupts motion prediction. *Cereb. Cortex* **25:** 1052–1059.
187. Wokke, M.E., A.R.E. Vandenbroucke, H.S. Scholte, *et al*. 2012. Confuse your illusion. *Psychol. Sci.* **24:** 63–71.
188. Strauss, M., J.D. Sitt, J.-R. King, *et al*. 2015. Disruption of hierarchical predictive coding during sleep. *Proc. Natl. Acad. Sci. USA* **112:** E1353–E1362.
189. Bekinschtein, T.A., S. Dehaene, B. Rohaut, *et al*. 2009. Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. USA* **106:** 1672–1677.
190. Chennu, S., V. Noreika, D. Gueorguiev, *et al*. 2013. Expectation and attention in hierarchical auditory prediction. *J. Neurosci.* **33:** 11194–11205.
191. Cornella, M., S. Leung, S. Grimm, *et al*. 2012. Detection of simple and pattern regularity violations occurs at different levels of the auditory hierarchy. *PLoS One* **7:** e43604.
192. Phillips, H.N., A. Blenkmann, L.E. Hughes, *et al*. 2016. Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. *Cortex* **82:** 192–205.
193. Uhrig, L., S. Dehaene & B. Jarraya. 2014. A hierarchy of responses to auditory regularities in the macaque brain. *J. Neurosci.* **34:** 1127–1132.
194. Kovács, G. & S.R. Schweinberger. 2016. Repetition suppression—an integrative view. *Cortex* **80:** 1–4.
195. Kumar, S., W. Sedley, K.V. Nourski, *et al*. 2011. Predictive coding and pitch processing in the auditory cortex. *J. Cogn. Neurosci.* **23:** 3084–3094.
196. Issa, E., C.F. Cadieu & J.J. DiCarlo. 2018. Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. *eLife* **7:** e42870.
197. Brandman, T. & M.V. Peelen. 2017. Interaction between scene and object processing revealed by human fMRI and MEG decoding. *J. Neurosci.* **37:** 7700–7710.
198. Harrison, L.M., K.E. Stephan, G. Rees, *et al*. 2007. Extra-classical receptive field effects measured in striate cortex with fMRI. *Neuroimage* **34:** 1199–1208.
199. Hsieh, P.J., E. Vul & N. Kanwisher. 2010. Recognition alters the spatial pattern of fMRI activation in early retinotopic cortex. *J. Neurophysiol.* **103:** 1501–1507.
200. Sterzer, P., J.-D. Haynes & G. Rees. 2006. Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT+/V5. *Neuroimage* **32:** 1308–1316.
201. Kok, P. & L.L.F. van Lieshout. 2016. Local expectation violations result in global activity gain in primary visual cortex. *Sci. Rep.* **6:** 2197.
202. Walsh, K. & D.P. McGovern. 2018. Expectation suppression dampens sensory representations of predicted stimuli. *J. Neurosci.* **38:** 10592–10594.
203. Revina, Y., L.S. Petro & L. Muckli. 2018. Cortical feedback signals generalise across different spatial frequencies of feedforward inputs. *Neuroimage* **180:** 280–290.
204. Gilbert, C.D. & W. Li. 2013. Top-down influences on visual processing. *Nat. Rev. Neurosci.* **14:** 350–363.
205. Budd, J.M.L. 1998. Extrastriate feedback to primary visual cortex in primates: a quantitative analysis of connectivity. *Proc. Biol. Sci.* **265:** 1037–1044.
206. Teufel, C. 2018. Prior object-knowledge sharpens properties of early visual feature-detectors. *Sci. Rep.* **8:** 400.

207. Spratling, M.W. 2010. Predictive coding as a model of response properties in cortical area V1. *J. Neurosci.* **30:** 3531–3543.
208. Spratling, M.W. 2013. Distinguishing theory from implementation in predictive coding accounts of brain function. *Behav. Brain Sci.* **36:** 231–232.
209. Sterzer, P., D.L. Sheinberg, P.C. Fletcher, *et al.* 2018. The predictive coding account of psychosis. *Biol. Psychiatry* **84:** 634–643.
210. Stefanics, G., J. Heinzle, A.A. Horváth, *et al.* 2018. Visual mismatch and predictive coding: a computational single-trial ERP study. *J. Neurosci.* **38:** 4020–4030.
211. Friston, K.J. & S. Kiebel. 2009. Attractors in song. *New Mathematics and Natural Computation* **5:** 1–32.
212. FitzGerald, T.H.B., R.J. Moran, K.J. Friston, *et al.* 2015. Precision and neuronal dynamics in the human posterior parietal cortex during evidence accumulation. *Neuroimage* **107:** 219–228.
213. Friston, K.J. 2018. Does predictive coding have a future? *Nat. Neurosci.* **21:** 1019–1021.
214. Lawrence, S.J., E. Formisano, L. Muckli, *et al.* 2019. Laminar fMRI: Applications for cognitive neuroscience. *Neuroimage* **197:** 785–791.

2.

The Role of Expectation in Binocular Rivalry

The fundamental question addressed in this thesis is how the brain resolves the inherent ambiguity in sensory stimulation, but centuries of research on a peculiar perceptual phenomenon has been driven by a fascination with what happens when it cannot. When the brain is confronted by the paradox of incompatible stimuli presented at corresponding retinal locations to each eye, it lapses into an unstable state characterised by spontaneous alternations in perception between the two conflicting images, such that one image dominates perception while the other is suppressed and a few seconds later the dominance reverses (Blake, 2001; Blake & Logothetis, 2002). This state, binocular rivalry, is considered a powerful tool for cognitive neuroscientists seeking to understand the neural mechanisms underlying perceptual awareness because it decouples conscious perceptual experience from retinal stimulation (Logothetis, 1998; Rees et al., 2002). While physically constant, the suppressed stimulus is entirely absent from the observer's awareness (Blake et al. 1998; Blake & Fox, 1974) and naive observers can exert little control over the contents of their experience at any given time (Blake, 1988; Meng & Tong, 2004). Although a unifying account of the underlying neural mechanisms of binocular rivalry has proven elusive (Brascamp et al., 2015; Tong et al., 2006), at a higher level, rivalry is thought of as the brain's ongoing effort to find a definite meaning with irresolvable evidence (Andrews & Purves, 1997; Blake, 2001). Some have proposed that prior expectations may shape these efforts (Hohwy et al., 2008). The purpose of this chapter is to empirically test this proposal.

The Mechanisms of Rivalry

Traditionally, rivalry dynamics are thought to emerge from a competition for perceptual awareness between mutually-inhibitory neural populations representing the discrepant interpretations of the retinal images (Alais, 2012; Kang & Blake, 2010; Tong, 2001). As these opposing populations are excited by the dichoptic images their mutual inhibition drives perception into one of two attractor states because a marginally stronger response in one population will inevitably lead to compounding inhibition and eventual dominance over the other (Lehky, 1988). It is assumed that this neural dynamic maps onto the perceptual experience of the emergence of a dominant percept and the suppression of the alternative. The 'rivalry' emerges when the dominant population succumbs to neural adaptation or fatigue, progressively weakening its suppression of the competing population until the dynamic shifts, and a new percept emerges (Alais, 2012; Leopold & Logothethis, 1999). To account for the stochastic nature of perceptual switches in rivalry, random noise is added to this process (Kang & Blake, 2010; Scocchia et al., 2014).

This account accords neatly with several empirical observations (Blake, 2001; Brascamp et al., 2015). For example, Kang and Blake (2010) physically removed the suppressed stimulus for a brief period immediately after the participant indicated that the other stimulus had become dominant, allowing adaptation to build for the dominant stimulus. They reported that the period of dominance when the suppressed stimulus was reintroduced decreased as the period of adaptation increased. Similarly, when a stimulus' dominance is artificially prolonged, by gradually enhancing its contrast, it exhibits increasingly brief dominance periods and when it is finally allowed to lapse into suppression, it remains suppressed for a longer period than usual (Blake, 2001; Blake et al., 1990). Lastly, sensitivity to contrast probes in the dominant stimulus decreases for longer periods of dominance, but increases with longer periods of suppression when the probe is in the suppressed stimulus (Alais et al., 2010). However, accumulating theoretical, psychophysical, and neurophysiological evidence suggests that adaptation is not a sufficient explanation for binocular rivalry (Alais, 2012; Leopold & Logothetis, 1999; Scocchia et al., 2014; Sterzer et al., 2009). For instance, the adaptation account predicts a strong correlation between the period for which a percept is suppressed and the duration of its subsequent dominance, since longer suppression allows for recovery from the effects of adaptation and therefore, more robust competition for dominance. However, efforts to test this prediction consistently find that successive durations of dominance are uncorrelated (Kang & Blake, 2010; Leopold & Logothetis, 1999; Walker, 1975).

Initially, rivalry was thought to represent interocular competition, where the mutually inhibitory populations described above were monocular channels in early visual areas (Tong et al., 2001). This was consistent with a number of psychophysical observations (Blake, 2001). For example, if the stimuli presented to each eye are swapped during rivalry, the perceived stimulus also immediately switches (Blake et al., 1980). However, other studies suggested that rivalry was a competition between perceptual representations at higher-levels, not monocular channels. Logothetis et al. (1996) found that when images are swapped between the eyes three times per second, far exceeding the rate of typical rivalry alternations, the percept of a single image persisted for several 'eye-swaps' at a time. Additionally, Kovacs et al. (1996) reported that binocular rivalry between two

images interwoven into patchworks, such that one image was the inverse composite of the other, produced coherent percepts of one or the other unified image. Ultimately, the purely eye-based competition account became untenable in the face of a wave of neurophysiological investigations that recast the conscious expression of binocular rivalry as a product of distributed perceptual reorganisation (Alais, 2012; Alais & Melcher, 2007; Wilson, 2003).

In monkeys, direct recordings from a variety of visual areas provided strong evidence that neither low- nor high-level explanations were individually sufficient. In early visual areas (e.g. LGN, V1, V2), neurons are largely unmoved by rivalry dynamics and fire in response to the physical stimulus, independent of its dominance or suppression (Lehky & Maunsell, 1996; Leopold & Logothetis, 1996; Logothetis & Schall, 1989; Wilke et al., 2009); in mid-level regions (e.g. V4, area MT, area MST) the responses of greater numbers of cells mimic the oscillations of perceptual content; finally, at higher-levels of the processing hierarchy (e.g. inferotemporal cortex, superior temporal sulcus), the majority of neurons respond in synchrony with the fluctuations of the subjectively perceived stimulus (Leopold & Logothetis, 1999; Sheinberg & Logothetis, 1997). Recently, Hesse and Tsao (2020) used a no-report paradigm to confirm that the behaviour of these neurons is not caused by the act of reporting the contents of perception. The response characteristics of these populations indicates that rivalry is neither the result of monocular gate-keeping or an event that takes place purely outside of sensory processing.

fMRI studies of humans also implicate both early and advanced stages of visual processing in binocular rivalry (Sterzer et al., 2009). Modulations of BOLD activity coincident with the vacillations of rivalry are most pronounced in higher-level areas (Blake et al., 2014; Zhang et al., 2017) and the magnitude of the response to the dominant stimulus in these regions is equivalent to the response to an unambiguous presentation of the same stimulus (Tong et al., 1998). The identity of both the dominant and suppressed stimulus can be predicted from BOLD activity in the fusiform face area and parahippocampal place area (Sterzer et al., 2008), suggesting that information about the suppressed stimulus does filter through to higher-levels of processing despite being excluded from perceptual awareness (Fang & He, 2005). However, in contrast to single-unit recordings, there is also evidence of modulations in BOLD and local field potential responses at earlier stages, including the thalamus (Haynes et al., 2005; Wilke et al., 2009; Wunderlich et al., 2005) and primary visual cortex (Polonsky et al., 2000; Tong & Engel, 2001), suggesting these regions are not ignorant of the changing perceptual states. In fact, the perceptual state can be decoded from BOLD activity in V1 during rivalry (Haynes & Rees, 2005) and transcranial magnetic stimulation of V1 can induce perceptual alternations (Pearson et al., 2007).

A limitation of the neuroimaging results is that they cannot distinguish whether this activity is caused by alternations or is causing alternations and this impedes efforts to adjudicate between feedforward and feedback models of binocular rivalry (Zhang et al., 2017). A number of studies have attempted to isolate the activity underlying perceptual alternations by comparing neural responses coinciding with real alternations to those associated with simulated alternations. These 'replay' conditions present an unambiguous stimulus that switches identity back and forth to recapitulate the timing of alternations reported in a previous rivalry trial, and

participants continue to report their perceptions to replicate the behaviour-related components of the neural response. This work indicated that, compared to stimulus-driven changes, spontaneous alternations are associated with identical responses in lower-level visual areas (e.g. V1) but stronger responses in extrastriate and parietofrontal regions (Lumer et al., 1998; Lumer & Rees, 1999). Sterzer and Kleinschmidt (2007) reported that the parietofrontal response to alternations preceded the response in lower-level areas, suggesting that the perceptual alternations may be initiated by higher-level processes. However, the validity of the 'replay' condition was subsequently criticised by some who pointed out that endogenous alternations are a more confusing and uncertain experience than the neat transitions that replay conditions portray (Blake et al., 2014). When Knapen et al. (2011) took additional measures to better mimic rivalry transitions in the replay condition, they found comparable activity in frontoparietal regions, suggesting that, instead of instigating the alternations, this activity may reflect the detection of conflict, task difficulty, the allocation of spatial attention, or the act of self-monitoring (Frassle et al., 2014).

Supplementing the psychophysical evidence that points to both feedforward and feedback contributions to rivalry (Alais, 2012; Blake, 2001; Scocchia et al., 2014), the neurophysiological evidence called for a hybrid of previous approaches. According to this hierarchical account, the sensory representation of the dominant stimulus is destabilised by accruing adaptation and the consequent conflict is signalled at lower-levels (e.g. Katyal et al., 2018; Said & Heeger, 2013), prompting higher-level areas to coordinate a resolution to the conflict via cortical feedback, resulting in a refreshed perceptual interpretation (Alais & Melcher, 2007; Leopold & Logothetis, 1999; Sterzer et al., 2009; Tong et al., 2006; Zhang et al., 2017). Some authors have ascribed this feedback influence to attention (e.g. Li et al., 2017), but others have developed accounts based on Bayesian inference (e.g. Gershman et al., 2012; Leptourgos et al., 2020).

Binocular Rivalry and Expectation

The predictive processing account of binocular rivalry incorporates many of the components of more traditional approaches, but reimagines them in terms of hierarchical prediction error minimisation (discussed in Chapter 1; Hohwy et al., 2008). Predictive feedback distributes predictions throughout the processing stream about likely patterns of incoming sensory stimulation, suppressing expected input so only discrepancies ascend to higher levels of processing in the form of prediction errors (Clark, 2013; Friston, 2005; Hohwy, 2012). According to this model, the dominant percept is the current perceptual hypothesis, which is determined by Bayesian principles as the best explanation for the current sensory evidence given prior experience. As the source of sensory predictions, input associated with the dominant percept is expected and suppressed, but failure to account for the other half of retinal stimulation produces prediction error. The current set of predictions cannot quell this error, since their source is excluded from perceptual awareness, so the only feedforward signals streaming through visual cortex represent the currently suppressed stimulus and call for a revision of the perceptual hypothesis to account for its presence. When the accumulation of prediction error

can no longer be ignored, the hypotheses are updated, perception alternates, and the dynamic reverses, with prediction error now advocating for the recently suppressed stimulus (Hohwy et al., 2008). This dynamic fits neatly with the empirical observation that the increase in dominance achieved by boosting the strength (e.g. contrast) of one of the rivalling stimuli is primarily driven by abbreviated suppression periods without much change to the duration of its dominance (Alais, 2012; Blake, 2001; Weilnhammer et al., 2017).

The predictive processing account is not specified at the level of neural implementation, so it is by claiming that binocular rivalry is a product of active cortical prediction that this model diverges from many of its peers. In discussions of the role of expectation in perception, it is often claimed that sensory expectations should be most influential when sensory input is most ambiguous (e.g. de Lange et al., 2018). Binocular rivalry is resolutely ambiguous, so it represents a promising testing ground for a theory that has proven difficult to assess empirically (Walsh et al., 2020). While the sensory input is normally held constant in binocular rivalry paradigms, predictive processing's account of rivalry can be tested by manipulating participants' priors. Numerous studies have provided evidence that prior experience influences the interpretation of ambiguous stimuli (Brascamp et al., 2007; Hsieh et al., 2010; Maloney et al., 2005; Panichello et al., 2013; Pearson & Brascamp, 2008; Pinto et al., 2015; Sterzer et al., 2008; Teufel et al., 2018). There are also some hints that long-standing priors in the form of environmental statistical regularities do influence the relative dominance of rivalling stimuli; natural images tend to dominate artificial images (Baker & Graf, 2009) and upright faces dominate inverted faces (Zhou et al., 2010). However, investigations of the influence of expectation in binocular rivalry have primarily focussed on immediate priming (e.g. Brascamp et al., 2007) or global context effects (e.g. Alais & Blake, 1999). Only a handful of studies have directly manipulated expectations on longer timescales that divorce higher-level 'expectation' from more immediate and local mechanisms.

In each of these studies of expectation in binocular rivalry, dominance is measured in terms of which stimulus is first perceived as rivalry onsets. When participants are presented with an apparent motion sequence of oriented gratings, the grating orientation that is predicted by the sequence dominates over a non-predicted orientation in subsequent rivalry (Andermane et al., 2020; Denison et al., 2011). When participants are misled to believe that a cue predicts the upcoming grating orientation, the cued grating dominates over the uncued grating (Andermane et al., 2020). Similarly, directional auditory 'sweeps' bias the perception of rivalrous random dot motion stimuli in the direction of the sound (Conrad et al., 2010). However, other studies have found the opposite relationship between dominance and expectation. Denison et al. (2016) trained participants to associate sequences of images in a statistical learning paradigm, but they found that unexpected images or image categories dominated rivalrous presentations with expected images or image categories. Finally, Chopin and Mamassian (2012) presented participants with a series of 1-4 oriented gratings interrupted by rivalrous gratings, requiring participants to indicate which grating they perceived. They found that a grating with a certain orientation was less likely to be perceived if that orientation was more frequent in the preceding series, which is a typical effect of adaptation (Thompson & Burr, 2009). However, the prevalence of that orientation in more distant stimulus history (5-13 minutes in the past) was positively correlated with the perceived orientation. They proposed that adaptation is a predictive mechanism that operates according to a kind of gambler's

fallacy: if the frequency of an orientation in recent history is disproportionate, the next stimulus is expected to have the alternative orientation so the established ratio can be restored.

In summary, proponents of predictive processing have pointed to binocular rivalry as a neat demonstration of prediction error minimisation at work (Hohwy et al., 2008), but relatively few empirical studies can speak to the veracity of these claims (Zhang et al., 2017). The study reported in this chapter aimed to test this account by instilling a prior expectation using an extended display of grating orientations with a statistical bias toward one of the competing orientations presented during a subsequent rivalry phase. Explicit cues manipulate attention as well as expectation (Summerfield & Egner 2009) and attention can influence the dynamics of rivalry (Chong & Blake, 2006; Hancock & Andrews, 2007). To take account of this, participants in the present study were asked to perform a target detection task at fixation during the prior-display, rendering the statistical bias task-irrelevant. In line with the predictive processing account, it was hypothesised that the statistical regularity would inform sensory predictions about upcoming input, such that perception would be initially biased toward the expected stimulus at rivalry onset. This prior was expected to erode over time when confronted with consistent and equivalent sensory evidence for both stimuli during rivalry. The study also sought to extend previous work by profiling the longevity of any expectation effect on more prolonged rivalry phases in addition to the analyses of the first reported percept after rivalry onsets.

Methodology

Participants

24 adults participated in this study. However, 4 failed to complete both sessions. A further 2 participants were rejected as their data suggested that they failed to experience rivalry consistently. Therefore, the analyses of rivalry dynamics were conducted on 18 participants (11 female, Range: 18-34, M = 22.6 ± 4.2). Four of these participants also participated in a follow-up experiment to measure adaptation to the test stimulus.

All participants reported normal or corrected-to-normal vision, no history of migraine or bad headaches, no history of epilepsy, and no sensitivity to flashing light. Participants provided informed written consent prior to testing and were paid a gratuity of €20 as compensation for their time. All procedures were approved by the Trinity College Dublin School of Psychology Ethics Committee and were in accordance with the Declaration of Helsinki.

Apparatus

The experiment was conducted in a dark, quiet room. Visual stimuli were generated using a custom PsychoPy experimental script and were presented on a 32 cm gamma-corrected CRT monitor (1024x768 resolution, 60 Hz frame rate) located in an isolated testing room in Trinity College Dublin. Binocular rivalry was induced using a custom-built mirror stereoscope, which was mounted on the edge of a desk, such that the viewing distance was 30 cm. The stereoscope facilitated separate and exclusive presentation of distinct grating stimuli to each of the participant's eyes. The stereoscope was composed of a chin rest, to comfortably stabilise the participant's head, and a set of four 50 mm diameter circular mirrors, each angled at 45° on the horizontal plane. A schematic of the experimental set up is provided in Figure 2.1.

Experimental Task

Rivalry Task

The experiment was composed of three phases: calibration, Baseline trials, and Test trials. Calibration was achieved by manually aligning two sets of nonius lines to create a coherent crosshair stimulus in the centre of the participant's field of vision (see Figure 2.1). The upper and right arms of the crosshairs were blue and were presented on the right side of the screen aligned with the right eye, while the lower and left arms were shown

in red and were presented on the left side of the screen aligned with the left eye. If presented together, the entire crosshair stimulus subtended 19°x19° of visual angle.

The Baseline and Test trials used coloured oriented grating stimuli presented at locations on the screen determined by the previous calibration phase. The Baseline trials were composed of a 'rivalry phase', which involved dichoptic presentation of these oriented gratings for 120 seconds while participants recorded the dominance durations of each percept as they experienced rivalry. The stimuli were annular gratings, oriented at ± 45° relative to vertical, with an outer radius of 18° visual angle and an inner radius of 3.58°, to accommodate the fixation point. The left-tilted grating was always presented to the left eye and the right-tilted grating was always presented to the right eye. The gratings were surrounded by matching annular fusion locks, which were designed to stabilise the rivalrous percepts. All grating stimuli were presented at 30% contrast with a spatial frequency of 0.34 cycles per degree. To reduce adaptation to the grating orientations, the phase of the grating was incremented on each frame at a rate of 10% of a cycle per second, giving the appearance that the phase continuously and smoothly drifted downwards and perpendicular to the grating tilt. To encourage rivalry, the gratings were either presented as having red and blue stripes or having green and purple stripes (Figure 2.2.A). The colour assigned to each grating was alternated between trials.

The Test trials included an additional element before the rivalry phase, a prior-instilling display was presented before measuring rivalry dynamics again. This prior-display was composed of a succession of rapidly changing (4 Hz) gratings with different orientations (20 possible orientations) and colours lasting 180 seconds in total. The statistics of this display were biased, such that in each testing session either the left-tilted or right-tilted grating from the Baseline trials comprised 25% of the gratings being presented. In the second testing session, the bias would favour the other grating tilt and the order of the sessions was counterbalanced across participants. All other gratings incorporated in the display were randomly chosen on each trial from a set of orientations from 0° (vertical) to 175° in steps of 5° . The colours were generated with three randomly selected RGB values from a set ranging from -1 to 1 in steps of 0.05. The order of the gratings in the display was pseudorandomly shuffled so that on no occasion was the same orientation presented contiguously. The display also incorporated a distractor task to limit awareness of the manipulation of orientation probability. The fixation dot, which was black throughout the experiment, would blink red between 3 and 9 times at random intervals, excluding the first and last 10% of the 180 second display, and participants were asked to count these targets. Importantly, both eyes were presented with identical stimuli during the prior-display. Only during the rivalry phase, were different stimuli presented to each eye. Immediately following the prior-display, a 16x16 pixelated mask, with a width of 30° of visual angle, was shown for 2 seconds to disrupt any afterimages. The colour of each pixel was randomly selected and was updated at 12 Hz. This was followed by a further 2 second inter-stimulus interval, where only the fusion lock was presented, before the rivalrous gratings were displayed and the rivalry phase began.

Tilt Aftereffect Task

A follow-up experiment using an orientation discrimination task was conducted to determine if the prior-display caused adaptation. Adaptation was expected to manifest as a tilt aftereffect, where the perception of vertical orientation is repelled away from the adapter orientation (McGovern et al., 2017; Thompson & Burr, 2009). This task used the same prior-display as the rivalry task, but the statistical bias towards only the right-tilted stimulus was manipulated across sessions. The task was composed of a baseline condition and three bias levels: 12.5%, 25%, and 50%, where in the 50% condition every second stimulus had a right-tilted orientation. Participants completed two sessions for each of the prior probability conditions in ascending order (completing the 50% condition in the final session), beginning with a baseline condition, where there was no prior-display and no top-up stimulus between test stimuli. Participants completed no more than two of the bias sessions in one day. One participant only completed one of the 12.5% bias sessions and another participant did not complete any 50% bias sessions and only completed one baseline session.

The gratings used in the prior-display and as test stimuli had all of the same properties as those used in the rivalry task, except all gratings were black and white. After exposure to the 180 second prior-display in each of the bias conditions, participants were asked to indicate if a series of gratings were tilted left or right from vertical. These test stimuli were composed of 20 sample gratings at each of seven possible orientations (-1.5°, -1°, -0.5°, 0°, 0.5°, 1°, 1.5°), where 0° refers to vertical. Each test stimulus was presented for 500 ms. Once the participant had responded, a 3 second 'top-up' prior-display was presented before the next stimulus. This top-up stimulus was identical to the prior-display apart from its duration. All of the stimuli were presented dichoptically, but the stimuli presented to each eye were always identical in this task.

Procedure

Rivalry Task

Each participant completed two sessions composed of two Baseline trials followed by five Test trials. The researcher first explained the concept of binocular rivalry and described the perceptual alternations that the participant would experience. The participant then sat in a comfortable chair and placed their chin in a chin rest in front of the mirror stereoscope. First, participants calibrated the stimuli on screen to ensure they were properly aligned with the mirrors. This calibration involved manually adjusting a crosshairs stimulus composed of two sets of differently coloured nonius lines presented on opposite sides of the screen, but perceived as a unified object after proper alignment (Figure 2.1.C). The participants were asked to make adjustments to the stimulus position on the screen using keyboard buttons until a fixation point in the middle of the crosshairs was perceived as centred in their field of view; these adjustments shifted both sets of lines equally in the same direction. They were then asked to adjust the stimulus' width until it was perceived as symmetric (see Figure 2.1.D); these adjustments shifted the sets of lines closer together or farther apart. When the participant reported that they were satisfied with their calibration, the full crosshairs stimulus was rapidly alternated

between the coordinates provided for the left stimulus and the right stimulus, it also changed colour as it alternated. The participant was told that if calibration was successful, the crosshairs would appear to flash in place and they may proceed to testing, but if there was any error in the alignment, they would appear to shift slightly with each change of colour and they should return to the calibration stages.

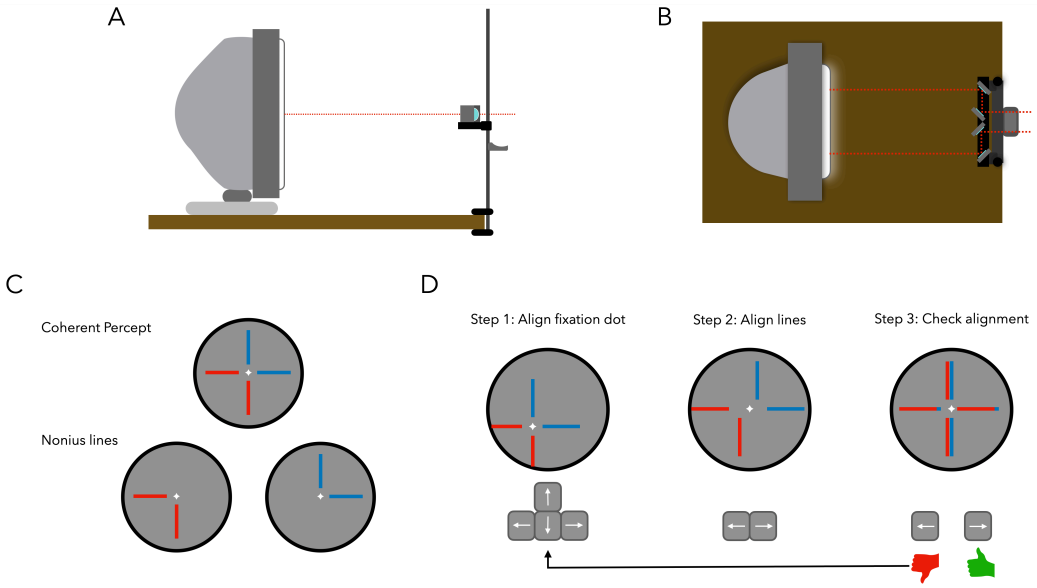


Figure 2.1. *A schematic of the apparatus and calibration stages of the rivalry experiment. A and B show how the mirror stereoscope was used to present dichotic stimuli. C) The nonius lines presented separately to each eye during the calibration phase and the coherent, unified percept that indicated correct alignment. D) The steps that participants progressed through to ensure that stimuli were correctly aligned. In the final step it can be seen that the stimulus is shifting left slightly as it changes from blue to red, so this participant would restart the calibration to address this error.*

Having completed the calibration, participants were told that in the next phase of the experiment a different grating would be presented to each eye and they would experience binocular rivalry. They were instructed to maintain fixation on the fixation dot throughout the experiment. They were asked to hold down the right arrow key when they perceived 80% or more of the grating that was tilted towards the right (45°) and hold down the left arrow key when they perceived 80% or more of the grating that was tilted towards the left (-45°). In periods when neither grating reached this level of dominance and their percept was mixed, they were instructed to not press any buttons. In this way, the dominant stimulus and the duration of each period of dominance was recorded. The researcher ensured that the participant understood this instruction by showing them a sample of each grating on a print out (see Figure 2.2.A). The rivalrous gratings were presented for 120 seconds and participants completed two of these Baseline trials at the beginning of each testing session.

Participants were told that they would now complete five trials with an added element (Figure 2.2.B). At the beginning of each trial participants would see a 180 second display, but they were instructed to ignore this display and focus on the fixation dot, which would blink red at random intervals. They were instructed to keep a count in their head of how many times the fixation dot blinked red. Participants were told that when this part of

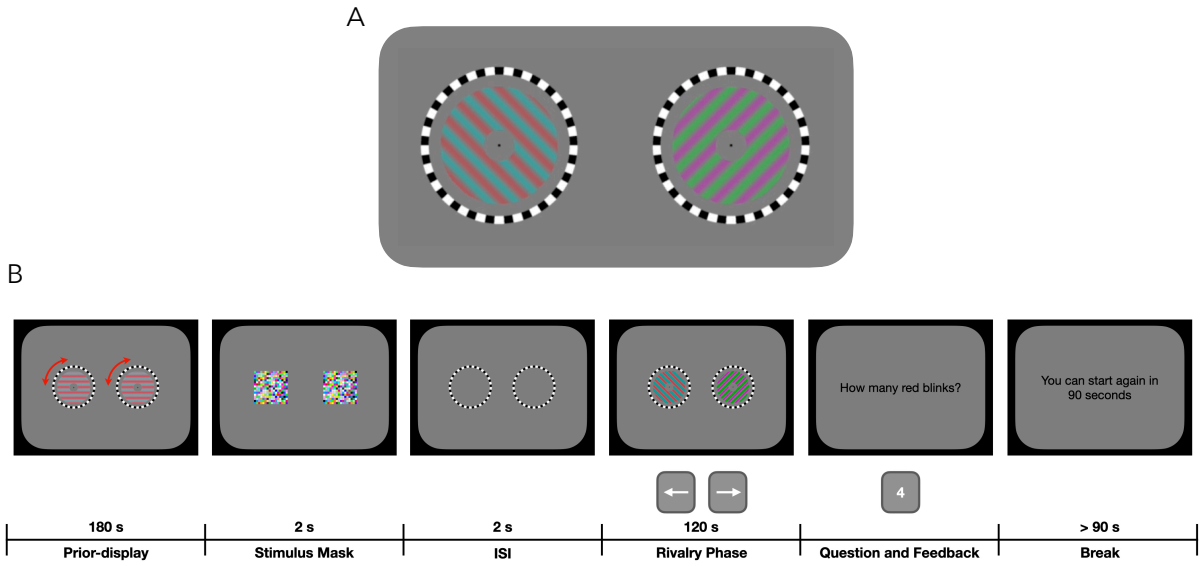


Figure 2.2. *A) The rivalrous gratings that were presented dichoptically during the rivalry phase of the Baseline and Test trials. The gratings were phase drifted to attenuate adaptation, the phase of the left grating drifted towards the bottom left corner and the phase of the right grating drifted towards the bottom right corner. The fusion lock surrounding the gratings was designed to stabilise the experience of rivalry. B) The task schematic for a Test trial. Baseline trials consisted of only the rivalry phase.*

the trial was over they would see a square made up of different coloured pixels for 2 seconds and then the 120 second rivalry phase that they had already experienced would begin again. Participants were told to respond in exactly the same fashion as they had previously. When this was completed, participants were asked how many times the fixation point had blinked red during the prior-display and received feedback on their response. A 90 second break was enforced between trials, but participants could choose to start the next trial any time after that interval.

Tilt Aftereffect Task

This experiment was conducted with an identical set up to the rivalry task and all participants that completed the tilt aftereffect task had already completed the rivalry task. Before all sessions, participants were reminded to maintain fixation at all times. Before each session began, participants completed the calibration, as they had done during the rivalry task. The first two sessions were baseline sessions, which only contained the orientation discrimination task. Participants were told that they would need to determine if each of a series of briefly presented gratings was tilted to the left or right and they were asked to use the left or right arrow key to indicate their response. They were informed that the task would wait for a response before moving on to the next grating stimulus. The remaining sessions incorporated the prior-display with varying levels of bias towards the right-tilted grating. Participants were told that these sessions would be similar to the baseline sessions, except they would be presented with the same rapidly-changing display that they had previously seen in the rivalry task. After this, they would begin the same orientation discrimination task that they had completed in the

baseline sessions. They were told that after each response the rapidly-changing display would reappear for a few seconds before the next grating stimulus was presented.

Statistical Analyses

The rivalry data were coded as the dominance time of the expected and unexpected grating, according to the bias in the prior-display. Extreme outliers, defined as values greater than three times the interquartile range from the median value, were removed in all analyses. As outliers, it was believed that these data points were unlikely to simply represent instances of stronger perceptual dominance generally and were more likely to be the result of inattention or fatigue. For completeness, the analyses were repeated with outliers included, but this did not meaningfully change any of the results. In all ANOVA analyses, a Greenhouse-Geiser correction was applied to the results when Mauchly's test indicated that the assumption of sphericity had been violated.

Rivalry Task

To produce a time course of rivalry dynamics, each trial was divided up into 2 second bins and the overall percentage dominance of each grating was calculated for each bin. For example, if a bin contained 1 second of dominance for the expected grating, 0.5 seconds of dominance for the unexpected grating, and 0.5 seconds of mixed percept (no buttons pressed), it would be recorded as 50% expected, 25% unexpected, and 25% mixed dominance. A 10 second rolling window was then applied to smooth this time course by averaging the data within the window and moving forward in 2 second steps. This process produced a time course that was used as an index of the temporal profile of rivalry dynamics.

To analyse the temporal nature of the effect, participants' time courses were divided into 8 bins. Since the bins were taken from the smoothed data, they do not correspond to precise moments during the rivalry phase, but 22 seconds of the original data contributed to each bin. The dominance percentage of the expected and unexpected stimuli on Baseline and Test trials was compared within each bin using a series of paired-samples t-tests. A false discovery rate (FDR) correction was applied, using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995), to account for the multiple comparisons. One extreme outlier was excluded from the Baseline data and one extreme outlier was excluded from the Test data.

Tilt Aftereffect Task

The data in the tilt aftereffect task were coded as the percentage of responses indicating a right-tilt at each of the test orientations and averaged within conditions for each participant. A Maximum Likelihood logistic psychometric function was fit to the averaged data for each condition, where parameters were estimated with parametric bootstrapping, using the Palamedes Psychophysics toolbox for Matlab (Prins & Kingdom, 2018). The point of subjective equality (PSE) was estimated for each participant for each condition using this psychometric model. The PSE is the model's estimate of the orientation of a test stimulus that would result in an equal number of responses indicating a rightward and leftward tilt (i.e. perceived vertical). A repeated-measures ANOVA was used to analyse the changes in PSE across conditions.

Results

It was hypothesised that there would be an initial effect of expectation on rivalry dynamics, which would be eroded over the course of the rivalry phase. This was first addressed by investigating the impact of expectation on the identity, onset time, and duration of the first percept that was reported at the beginning of each Test rivalry phase (Figure 2.3). A one-sample t-test indicated that the unexpected grating was significantly more likely to be reported as the first percept after the prior-display ($t(17) = 3$, $p = 0.0081$). The onset and duration data were not normally distributed, so they were analysed with a non-parametric Wilcoxon Signed Ranks test. There was no significant difference between the ranks of the onset times of expected and unexpected percepts ($Z = -1.459$, $p = 0.145$). Expectation also did not have a significant impact on the durations of the first percepts ($Z = -1.459$, $p = 0.145$).

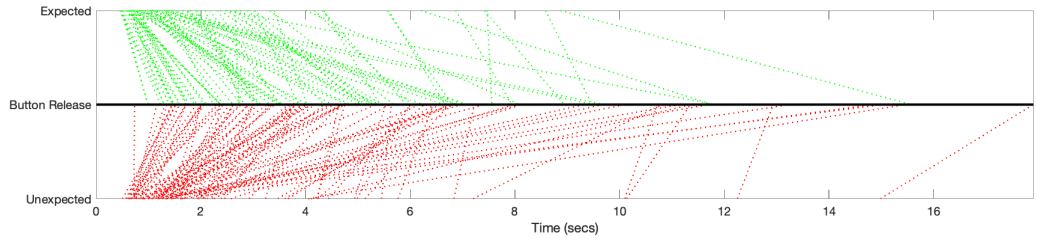


Figure 2.3. *Each recorded 'first percept' is plotted as green (expected percept) or red (unexpected percept). The point that each trace emerges from the x-axis marks the onset time of that percept. The duration of the percept is represented by the distance between its emergence from the x-axis and it reaching the line representing button release (i.e. the participant reporting that the percept was no longer dominant).*

Next, an analysis of the binned data investigated the possibility that there was an overall bias in favour of the expected stimulus in the early stages of the rivalry phase, which diminished over time (Figure 2.4). The Baseline and Test condition time courses were divided into 8 bins and paired samples t-tests compared the percentage of the bin dominated by the expected stimulus to the percentage of the bin dominated by the unexpected stimulus (referred to as 'marginal expected dominance'). Tables 1 and 2 show the results of these t-tests, with an FDR correction for multiple comparisons. Unsurprisingly, there was no significant difference in the dominance of expected and unexpected stimuli in any of the Baseline bins. However, in the Test trials, the unexpected stimulus was significantly more dominant than the expected stimulus in Bin 1 ($t(17) = -4.153$, $p = 0.008$) and Bin 2 ($t(17) = -2.849$, $p = 0.044$) , but there was no significant difference in the remaining Test bins. As the bins partitioned a time course that had been smoothed by a rolling window (see Statistical Analyses), they do not correspond to precise moments in the rivalry phase, but the first 36 seconds of data contributed to Bin 1 and Bin 2.

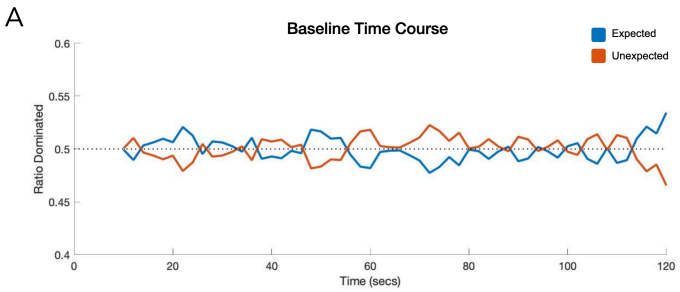
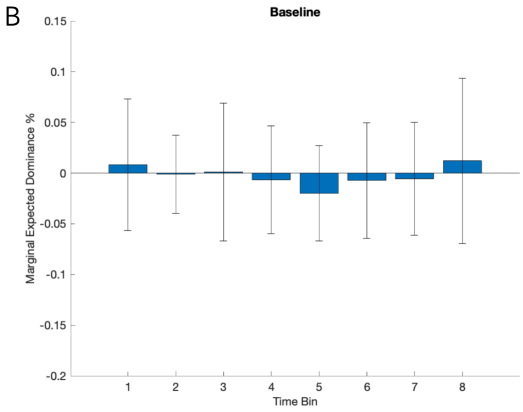


Table 1. Binned Marginal Expected Dominance in the Baseline Condition

Baseline	Mean	SD	t	df	Sig.
Bin 1	0.008	0.065	0.540	17	0.913
Bin 2	-0.003	0.040	-0.259	16	0.913
Bin 3	0.001	0.068	0.066	17	0.948
Bin 4	-0.005	0.054	-0.429	17	0.913
Bin 5	-0.019	0.046	-1.738	17	0.8
Bin 6	-0.007	0.057	-0.551	17	0.913
Bin 7	-0.005	0.056	-0.410	17	0.913
Bin 8	0.013	0.082	0.685	17	0.913



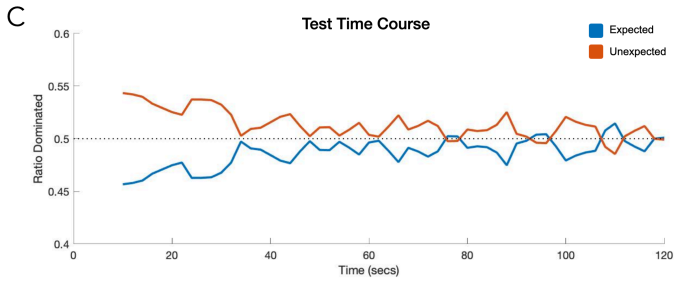


Table 2. Binned Marginal Expected Dominance in the Test Condition

Test	Mean	SD	t	df	Sig.
Bin 1	-0.050	0.052	-4.153	17	**0.008**
Bin 2	-0.040	0.059	-2.849	17	**0.044**
Bin 3	-0.021	0.057	-1.610	17	0.299
Bin 4	-0.013	0.056	-0.941	17	0.366
Bin 5	-0.015	0.051	-1.275	17	0.299
Bin 6	-0.016	0.049	-1.405	17	0.299
Bin 7	-0.014	0.064	-0.928	17	0.366
Bin 8	-0.012	0.039	-1.266	16	0.299

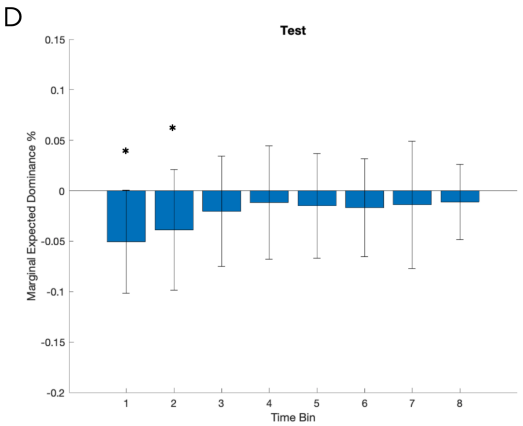


Figure 2.4. *The results of the binned analysis. A & C) The time course of dominance for expected and unexpected stimuli, with mixed percept time removed. Tables 1 & 2 report the results of the paired samples t-tests. The binned plots (B & D) refer to the 'Marginal Expected Dominance', which is the percentage of the dominance time in each bin associated with the expected stimulus minus the percentage of dominance time associated with the unexpected stimulus. For comparison to the Test trials, the stimuli are also labelled 'expected' and 'unexpected' on Baseline trials. This is an arbitrary mapping to provide a control condition. There is no expectation on Baseline trials, so 'expected' and 'unexpected' here refer to the expectation that would be set on subsequent Test trials in that testing session.*

A follow-up regression analysis was used to better characterise the temporal profile of this effect. The regression analyses were carried out on the difference between expected and unexpected dominance rates across the 8 bins used in the previous analysis. The simple regression model based on Time Bin did not account for a significant amount of the variance in rivalry dynamics for the Baseline trials ($R^2 = 0$, $F(1,141) = 0.011$, $p = 0.917$), but was significant for Test trials ($R^2 = 0.043$, $F(1,141) = 6.367$, $p = 0.013$). The model based on the Test trials suggests that there is initially a 4.5% bias towards the unexpected stimulus, which reduces by 0.5% for each consecutive time bin (Figure 2.5).

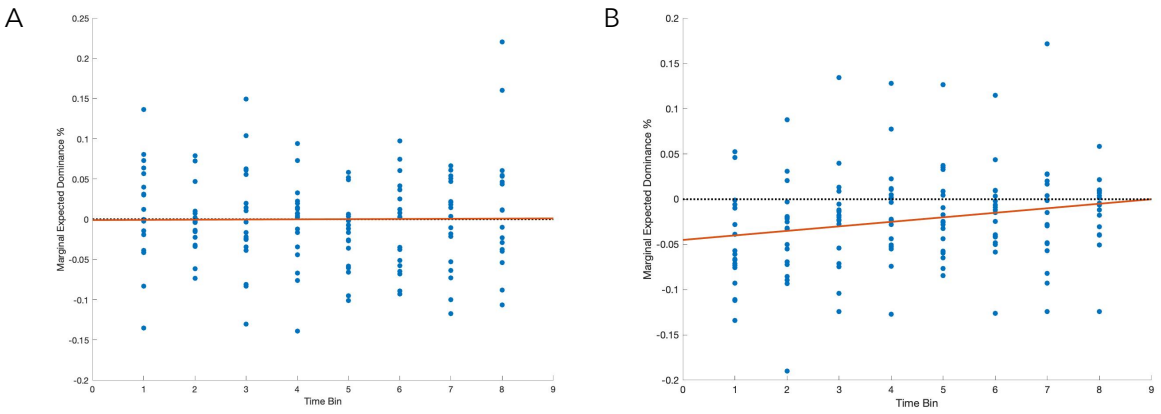


Figure 2.5. *Scatterplots showing the model fit for the baseline (A) and test (B) regression analyses. Each data point represents the difference between the percentage of the bin dominated by the expected and unexpected stimuli for one participant.*

Since the statistical regularities of the prior-display were intended to instil an expectation, it was hypothesised that the magnitude of any effect might increase over the course of the five exposures within each session as the precision of sensory expectations increases (Figure 2.6). This question was addressed in a factorial repeated-measures ANOVA using Bin Number and Trial Number as independent variables. The ANOVA confirmed that there was a significant main effect of the Bin Number on the difference between the dominance rate of the

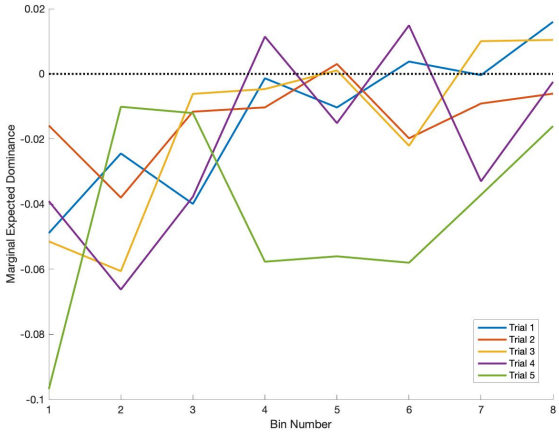


Figure 2.6. *Rivalry dynamics across repeated exposures to the prior-display. The y-axis represents the difference between the expected and unexpected dominance rates for each bin.*

expected and unexpected stimuli ($F(3.698,62.861) = 3.362$, $p = 0.017$), but there was no main effect of Trial Number ($F(2.215,37.653) = 0.674$, $p = 0.530$) and no interaction between Bin Number and Trial Number ($F(8.143,138.43) = 1.035$, $p = 0.413$).

Tilt Aftereffect

A separate dataset was collected to investigate if adaptation to the grating orientation presented more frequently in the prior-display might account for the effect observed in the above analyses. Participants completed an orientation discrimination task using the method of constant stimuli, following exposure to the prior-display with varying levels of bias towards the right-tilted grating (0%, 12.5%, 25%, 50%). Psychometric functions were estimated from participant responses to determine if the display induced a tilt-aftereffect. These psychometric functions for each participant and each condition are shown in Figure 2.7. Adaptation would be expected to shift participants' point of subjective equality (PSE) in the positive direction with increasingly biased displays, indicating that a greater rightward-tilt is required to perceive the grating as vertically oriented. However, this pattern was only evident in one participant's data (yellow) and a repeated-measures ANOVA indicated that there was no effect of the bias in the prior-display on tilt discrimination ($F(3,6) = 0.583$, $p = 0.647$). A Bayesian analysis ($BF = 0.524$) indicated that this was anecdotal evidence of there being no tilt aftereffect.

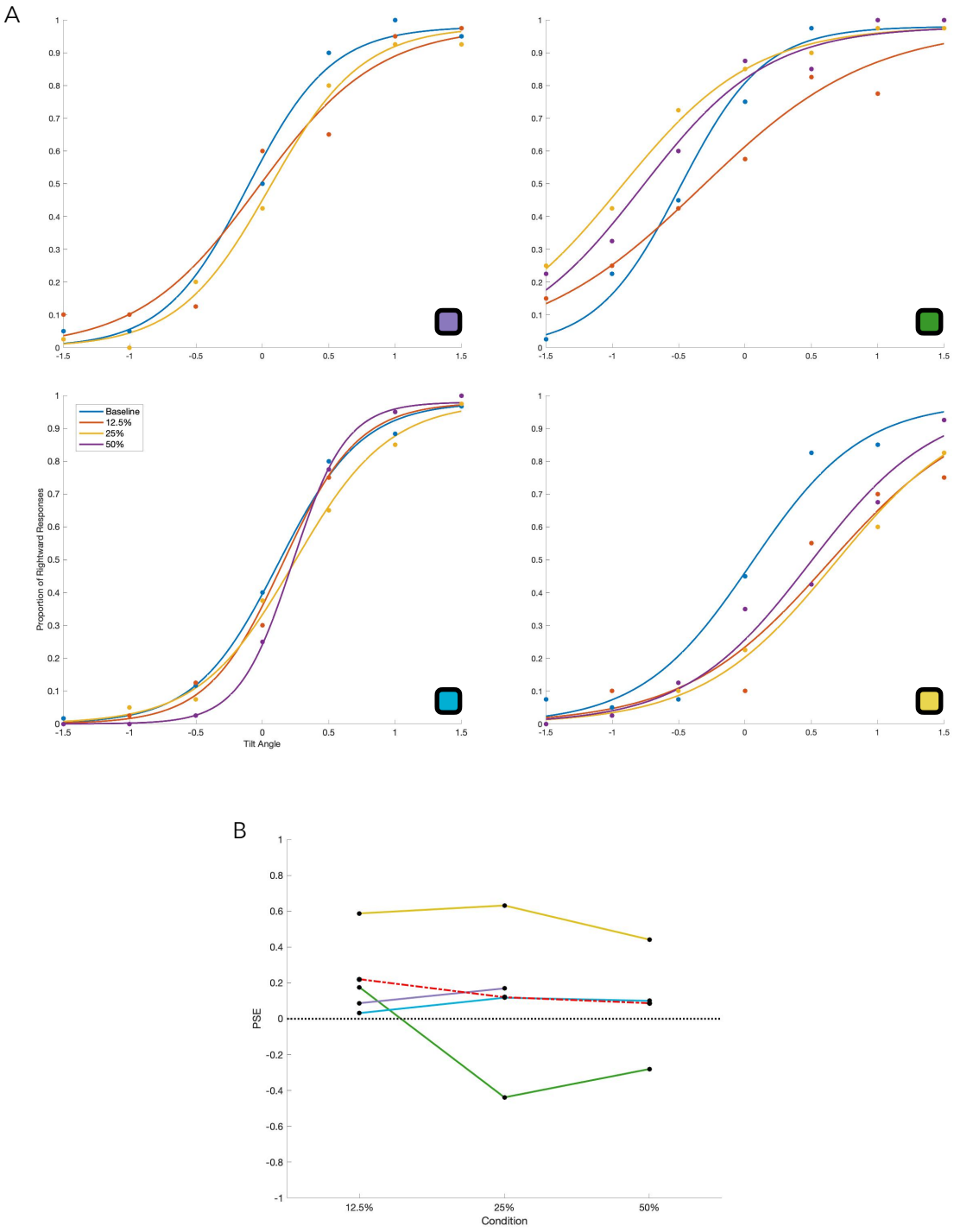


Figure 2.7. *The results of the tilt aftereffect study. A) The Psychometric functions for each of the participants across all levels of bias tested. Each participant is colour-coded so they can be identified on the PSE plot below. B) The changes in participants' baselined PSEs across conditions. The broken red line represents the mean across all participants.*

Discussion

This study sought to test the predictive processing account of binocular rivalry by determining if exposure to statistical regularities biases subsequent rivalry dynamics, in line with the predictive processing contention that the perceptual awareness of ambiguous stimuli is biased towards prior expectations. However, participants exhibited the opposite effect: the unexpected grating was more likely to be reported as the first percept as rivalry onset and it continued to dominate the early phase of rivalry. A follow-up study suggested that the prior-display did not cause any systematic change in the perception of orientation, suggesting the effect may not be attributable to adaptation.

Few studies have investigated the effects of expectation on rivalry dynamics, but there is precedent for this repulsive effect. For example, Denison et al. (2016) presented participants with a binocular stream of images composed of learned sequences. They found that images in unexpected sequence positions were more likely to be immediately perceived when presented in rivalry with the image that was expected in that sequence position. Instead of priors weighting perceptual hypotheses in rivalry, they suggested that perceptual awareness is drawn to the more surprising stimulus, which may be more informative in updating beliefs about the state of the environment. Several monkey neurophysiology studies using similar paradigms have reported suppressed responses to expected images in single-unit recordings of inferotemporal cortex (Kaposvari et al., 2018; Meyer et al., 2014; Meyer & Olson, 2011; Ramachandran et al., 2016) and the activity of the same neural populations has been found to reflect the subjective percept, not the physical stimulus, during rivalry (Hesse & Tsao, 2020; Leopold & Logothetis, 1999; Sheinberg & Logothetis, 1997). Together, this suggests that there may be an association between the suppression of neural responses to expected stimuli and the perceptual suppression of expected stimuli in rivalry.

Consistent with Denison et al., Chopin and Mamassian (2012) reported that perceptual awareness during rivalry was negatively correlated with the prevalence of stimulus orientations in recent stimulus history, but they also found a positive correlation with orientation prevalence in more distant stimulus history. They proposed a model that explained this effect as the product of 'predictive adaptation' reflecting an expectation that short-term fluctuations in stimulus history will regress towards the average statistics of long-term stimulus history. While their estimated three minute window for the repulsive effect accords neatly with the repulsive effect in the present study following the three minute prior-display, the estimated window for the attractive bias (approximately 5-13 mins) extends beyond the duration of any of the individual trials in this study. It might be expected that this attractive bias should emerge over the course of several trials from the cumulative exposure to the prior-display, but no such effect was observed. It's not clear that the exact same effects should be expected in the present dataset because Chopin and Mamassian maintained an overall balance between the number of left and right-tilted gratings in the stimulus stream across the experiment, whereas in the present study, the statistics of the prior-display were consistently biased towards the expected grating at all times.

Adaptation to a grating stimulus with a particular orientation depresses the responses of neurons tuned to that orientation (Thompson & Burr, 2009). When a new grating with a slightly different orientation is presented, the collective distribution of responses from orientation-selective neurons is skewed away from the adapting orientation, leading to a perceptual exaggeration of the tilt away from the adaptor orientation (He & MacLeod, 2001). This 'tilt aftereffect' accumulates over longer exposure to the adaptor stimulus (Patterson et al., 2013), so one interpretation of the present results is that they are unrelated to expectation and simply reflect adaptation to the more frequent stimulus during the 3 minute prior-display. To test this explanation, a follow-up study measuring orientation discrimination after exposure to the prior-display was conducted. However, the experiment provided no evidence that the repulsion effect was caused by adaptation to the expected stimulus. This result contrasts with Chopin and Mamassian who found that their stimuli did produce tilt aftereffects, with more recent concentrations of a particular orientation having a repulsive effect on discriminations and longer-term concentrations having an attractive effect. However, unlike the prior-display, which contained twenty orientations, Chopin and Mamassian's stimulus stream was only made up of two orientations, which likely strengthened the manifestation of adaptation in that study. There are several additional differences that complicate efforts to compare the present findings with those of Chopin and Mamassian. In the present study, participants were exposed to a 3 minute prior-display before 2 minutes of continuous rivalry and then at least a 90 s break, whereas in the study by Chopin and Mamassian, there were short series of 1-4 gratings interrupted by a transient rivalry phase requiring a response, but otherwise presented in a continuous stream that far exceeded the duration of a single trial in the present study. Additionally, rivalry onset 200 ms after the previous grating in Chopin and Mamassian's paradigm, whereas here, there was a 2 second mask stimulus and another 2 second inter-stimulus interval. The longer gaps to rivalry onset implemented in the present study are likely to have influenced the manifestation of any adaptation effect, as adaptation is known to decay quickly with recovery time (Greenlee et al., 1991; Kohn, 2007).

The prior-display used in the present study is not typical of the stimuli used to test for adaptation effects. Adaptation studies normally use prolonged exposure to static adaptors, but the present study used a dynamic display with a statistical bias and tested its effects with rivalrous gratings that were phase modulated to minimise the effects of adaptation. To our knowledge, other than the study by Chopin and Mamassian, only one psychophysics investigation of orientation adaptation used a comparably dynamic stimulus. Gekas et al. (2019) presented participants with a series of briefly shown Gabor patches with a statistical bias towards a particular orientation. They reported a clear tilt aftereffect, with perceptual judgements exhibiting a repulsive effect from the more common orientation, and their modelling suggested that this effect arose from several minutes of stimulus history. Given these results, it is somewhat surprising that we found no consistent trends across participants hinting at the role of adaptation and only one participant exhibited a tilt aftereffect. The absence of the phase modulation for the test stimuli and the distinction between the prolonged prior-display and stimulus stream interrupted by regular perceptual decisions may be influential factors in the difference between Chopin and Mamassian's tilt aftereffect and the present results. However, an important limitation of the present study is that the sample used in the tilt aftereffect experiment was very small. Although small samples are not unusual in psychophysics research, this is typically compensated for with extended testing, often of highly-experienced

participants. Participants in the tilt aftereffect experiment underwent a more modest testing regime, so it is not clear whether the sample size and number of testing sessions were simply too limited to detect an effect.

Apart from the tilt aftereffect, a second consequence of orientation-specific adaptation is a reduction in the sensitivity of neurons tuned to the adapting orientation. This 'orientation-selective contrast adaptation' reduces the effective contrast of stimuli that share the adapted orientation, such that they need to be presented at higher contrast levels to be detected after adaptation (Blakemore et al., 1973; He & MacLeod, 2001). It is possible that the diversity of orientations presented in the prior-display prevented the development of a robust tilt aftereffect, since the tilt aftereffect is associated with the population response of orientation-selective neurons. However, orientation-selective contrast adaptation is unlikely to have been attenuated in the same way, which suggests that the effective contrast of the expected orientation may have been reduced when the exact same orientation was presented during rivalry. This may explain the result because higher contrast stimuli dominate lower contrast stimuli in rivalry (Blake, 2001). Indeed, orientation-selective contrast adaptation has been shown to weaken an steady-state visual evoked potential (SSVEP) associated with a stimulus sharing the adapted orientation (Vergeer et al., 2018) and the SSVEP has also been shown to reliably track the dynamics of rivalry alternations, with reduced amplitudes for the suppressed stimulus (Brown & Norcia, 1997). A future study could collect data from a larger sample testing for both the tilt aftereffect and orientation-selective contrast adaptation.

While it would be useful to more definitively determine the contribution of adaptation to the current results, the detection of adaptation would not constitute evidence against predictive processing per se. The predictive processing account of binocular rivalry is framed as 'epistemological' by Hohwy et al. (2008), specifically because it does not offer a novel implementation of the underlying neural circuitry, but an algorithmic explanation of why perception behaves as it does by situating rivalry within a unifying framework of prediction error minimisation. In focussing on the higher-level explanation for rivalry dynamics, the authors assume that the neural instantiation of the dynamics described in the predictive processing account involve adaptation. For example, Hohwy et al. (2008) suggest that the process of prediction error from the currently suppressed stimulus calling for a revision of perceptual hypotheses would be mediated by adaptation of the neuronal representation of the currently dominant stimulus. From this perspective, adaptation is just another mechanism within the predictive processing framework, which is expected to operate according to the overarching objective of minimising prediction error. Thus, the observation that perceptual awareness was biased away from prior expectations conflicts with predictive processing accounts whether it is underpinned by adaptation or not.

References

Alais, D., & Blake, R. (1999). Grouping visual features during binocular rivalry. *Vision Research*, *39*, 4341–4353.

Alais, D., Cass, J., O'Shea, R. P., & Blake, R. (2010). Visual sensitivity underlying changes in visual consciousness. *Current Biology*, *20*, 1362–1367.

Alais, D., & Melcher, D. (2007). Strength and coherence of binocular rivalry depends on shared stimulus complexity. *Vision Research*, *47*, 269–279.

Alais, D. (2012). Binocular rivalry: competition and inhibition in visual perception. *Wiley Interdisciplinary Reviews. Cognitive Science*, *3*, 87–103.

Andermane, N., Bosten, J., Seth, A., & Ward, J. (2020). Individual differences in the tendency to see the expected. *Cognition and Consciousness, 85*, 102989.

Andrews, T. J., & Purves, D. (1997). Similarities in normal and binocularly rivalrous viewing. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 9905–9908.

Baker, D. H., & Graf, E. W. (2009). Natural images dominate in binocular rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 5436–5441.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 289–300.

Blakemore, C., Muncey, J. P., & Ridley, R. M. (1973). Stimulus specificity in the human visual system. *Vision Research*, *13*, 1915–1931.

Blake, R., & Fox, R. (1974). Binocular rivalry suppression: insensitive to spatial frequency and orientation change. *Vision Research*, *14*, 687–692.

Blake, R., Westendorf, D., & Fox, R. (1990). Temporal perturbations of binocular rivalry. *Perception & Psychophysics*, *48*, 593–602.

Blake, R., Westendorf, D. H., & Overton, R. (1980). What is suppressed during binocular rivalry? *Perception*, *9*, 223–231.

Blake, R., Yu, K., Lokey, M., & Norman, H. (1998). Binocular rivalry and motion perception. *Journal of Cognitive Neuroscience*, *10*, 46–60.

Blake, R., Brascamp, J., & Heeger, D. J. (2014). Can binocular rivalry reveal neural correlates of consciousness? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 20130211.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews. Neuroscience*, *3*, 13–21.

Blake, R. (2001). A Primer on Binocular Rivalry, Including Current Controversies. *Brain & Mind*, 5–38.

Blake, R. (1988). Dichoptic reading: the role of meaning in binocular rivalry. *Perception & Psychophysics*, *44*, 133–141.

Brascamp, J. W., Knapen, T. H. J., Kanai, R., van Ee, R., & van den Berg, A. V. (2007). Flash suppression and flash facilitation in binocular rivalry. *Journal of Vision*, *7*, 12.1-12.

Brascamp, J. W., Klink, P. C., & Levelt, W. J. M. (2015). The "laws" of binocular rivalry: 50 years of Levelt's propositions. *Vision Research*, *109*, 20–37.

Brown, R. J., & Norcia, A. M. (1997). A method for investigating binocular rivalry in real-time with the steady-state VEP. *Vision Research*, *37*, 2401–2408.

Chong, S. C., & Blake, R. (2006). Exogenous attention and endogenous attention influence initial dominance in binocular rivalry. *Vision Research*, *46*, 1794–1803.

Chopin, A., & Mamassian, P. (2012). Predictive properties of visual adaptation. *Current Biology*, *22*, 622–626.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204.

Conrad, V., Bartels, A., Kleiner, M., & Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision*, *10*, 27.

Denison, R. N., Piazza, E. A., & Silver, M. A. (2011). Predictive Context Influences Perceptual Selection during Binocular Rivalry. *Frontiers in Human Neuroscience*, *5*, 166.

Denison, R. N., Sheynin, J., & Silver, M. A. (2016). Perceptual suppression of predicted natural images. *Journal of Vision*, *16*, 6.

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*, 764–779.

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neuroscience*, *8*, 1380–1385.

Frässle, S., Sommer, J., Jansen, A., Naber, M., & Einhäuser, W. (2014). Binocular rivalry: frontal activity relates to introspection and action but not to perception. *The Journal of Neuroscience*, *34*, 1738–1747.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*, 815–836.

Gekas, N., McDermott, K. C., & Mamassian, P. (2019). Disambiguating serial effects of multiple timescales. *Journal of Vision*, *19*, 24.

Gershman, S. J., Vul, E., & Tenenbaum, J. B. (2012). Multistability and perceptual inference. *Neural Computation*, *24*, 1–24.

Greenlee, M. W., Georgeson, M. A., Magnussen, S., & Harris, J. P. (1991). The time course of adaptation to spatial contrast. *Vision Research*, *31*, 223–236.

Hancock, S., & Andrews, T. J. (2007). The role of voluntary and involuntary attention in selecting perceptual dominance during binocular rivalry. *Perception*, *36*, 288–298.

Haynes, J.-D., Deichmann, R., & Rees, G. (2005). Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. *Nature*, *438*, 496–499.

Haynes, J.-D., & Rees, G. (2005). Predicting the stream of consciousness from activity in human visual cortex. *Current Biology*, *15*, 1301–1307.

Hesse, J. K., & Tsao, D. Y. (2020). A new no-report paradigm reveals that face cells encode both consciously perceived and suppressed stimuli. ELife, 9.

He, S., & MacLeod, D. I. (2001). Orientation-selective adaptation and tilt after-effect from invisible patterns. *Nature*, *411*, 473–476.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition*, *108*, 687–701.

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.

Hsieh, P. J., Vul, E., & Kanwisher, N. (2010). Recognition alters the spatial pattern of FMRI activation in early retinotopic cortex. *Journal of Neurophysiology*, *103*, 1501–1507.

Kang, M.-S., & Blake, R. (2010). What causes alternations in dominance during binocular rivalry? *Attention, Perception & Psychophysics*, *72*, 179–186.

Kaposvari, P., Kumar, S., & Vogels, R. (2018). Statistical learning signals in macaque inferior temporal cortex. *Cerebral Cortex*, *28*, 250–266.

Katyal, S., Vergeer, M., He, S., He, B., & Engel, S. A. (2018). Conflict-sensitive neurons gate interocular suppression in human visual cortex. *Scientific Reports*, *8*, 1239.

Knapen, T., Brascamp, J., Pearson, J., van Ee, R., & Blake, R. (2011). The role of frontal and parietal brain areas in bistable perception. *The Journal of Neuroscience*, *31*, 10293–10301.

Kohn, A. (2007). Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, *97*, 3155–3164.

Kovács, I., Papathomas, T. V., Yang, M., & Fehér, A. (1996). When the brain changes its mind: interocular grouping during binocular rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 15508–15511.

Lehky, S. R., & Maunsell, J. H. (1996). No binocular rivalry in the LGN of alert macaque monkeys. *Vision Research*, *36*, 1225–1234.

Lehky, S. R. (1988). An astable multivibrator model of binocular rivalry. *Perception*, *17*, 215–228.

Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, *379*, 549–553.

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences*, *3*, 254–264.

Leptourgos, P., Bouttier, V., Jardri, R., & Denève, S. (2020). A functional theory of bistable perception based on dynamical circular inference. *PLoS Computational Biology*, *16*, e1008480.

Li, H.-H., Rankin, J., Rinzel, J., Carrasco, M., & Heeger, D. J. (2017). Attention model of binocular rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, E6192–E6201.

Logothetis, N. K., Leopold, D. A., & Sheinberg, D. L. (1996). What is rivalling during binocular rivalry? *Nature*, *380*, 621–624.

Logothetis, N. K., & Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science*, *245*, 761–763.

Logothetis, N. K. (1998). Single units and conscious vision. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*, 1801–1818.

Lumer, E. D., Friston, K. J., & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science*, *280*, 1930–1934.

Lumer, E. D., & Rees, G. (1999). Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 1669–1673.

Maloney, L. T., Dal Martello, M. F., Sahm, C., & Spillmann, L. (2005). Past trials influence perception of ambiguous motion quartets through pattern completion. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 3164–3169.

McGovern, D. P., Walsh, K. S., Bell, J., & Newell, F. N. (2017). Individual differences in context-dependent effects reveal common mechanisms underlying the direction aftereffect and direction repulsion. *Vision Research*, *141*, 109–116.

Meng, M., & Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of Vision*, *4*, 539–551.

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 19401–19406.

Meyer, T., Ramachandran, S., & Olson, C. R. (2014). Statistical learning of serial visual transitions by neurons in monkey inferotemporal cortex. *The Journal of Neuroscience*, *34*, 9332–9337.

Panichello, M. F., Cheung, O. S., & Bar, M. (2012). Predictive feedback and conscious visual experience. *Frontiers in Psychology*, *3*, 620.

Patterson, C. A., Wissig, S. C., & Kohn, A. (2013). Distinct effects of brief and prolonged adaptation on orientation tuning in primary visual cortex. *The Journal of Neuroscience*, *33*, 532–543.

Pearson, J., & Brascamp, J. (2008). Sensory memory for ambiguous vision. *Trends in Cognitive Sciences*, *12*, 334–341.

Pearson, J., Tadin, D., & Blake, R. (2007). The effects of transcranial magnetic stimulation on visual rivalry. *Journal of Vision*, *7*, 2.1-11.

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*, 13.

Polonsky, A., Blake, R., Braun, J., & Heeger, D. J. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nature Neuroscience*, *3*, 1153–1159.

Prins, N., & Kingdom, F. A. A. (2018). Applying the Model-Comparison Approach to Test Specific Research Hypotheses in Psychophysical Research Using the Palamedes Toolbox. *Frontiers in Psychology*, *9*, 1250.

Ramachandran, S., Meyer, T., & Olson, C. R. (2016). Prediction suppression in monkey inferotemporal cortex depends on the conditional probability between images. *Journal of Neurophysiology*, *115*, 355–362.

Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. *Nature Reviews. Neuroscience*, *3*, 261–270.

Said, C. P., & Heeger, D. J. (2013). A model of binocular rivalry and cross-orientation suppression. *PLoS Computational Biology*, *9*, e1002991.

Scocchia, L., Valsecchi, M., & Triesch, J. (2014). Top-down influences on ambiguous perception: the role of stable and transient states of the observer. *Frontiers in Human Neuroscience*, *8*, 979.

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 3408–3413.

Sterzer, P., Frith, C., & Petrovic, P. (2008). Believing is seeing: expectations alter visual awareness. *Current Biology*, *18*, R697-8.

Sterzer, P., Haynes, J.-D., & Rees, G. (2008). Fine-scale activity patterns in high-level visual areas encode the category of invisible objects. *Journal of Vision*, *8*, 10.1-12.

Sterzer, P., Kleinschmidt, A., & Rees, G. (2009). The neural bases of multistable perception. *Trends in Cognitive Sciences*, *13*, 310–318.

Sterzer, P., & Kleinschmidt, A. (2007). A neural basis for inference in perceptual ambiguity. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 323–328.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*, 403–409.

Teufel, C., Dakin, S. C., & Fletcher, P. C. (2018). Prior object-knowledge sharpens properties of early visual feature-detectors. *Scientific Reports*, *8*, 10853.

Thompson, P., & Burr, D. (2009). Visual aftereffects. *Current Biology*, *19*, R11-4.

Tong, F., & Engel, S. A. (2001). Interocular rivalry revealed in the human cortical blind-spot representation. *Nature*, *411*, 195–199.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, *21*, 753–759.

Tong, F., Meng, M., & Blake, R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences*, *10*, 502–511.

Tong, F.. (2001). Competing Theories of Binocular Rivalry: A Possible Resolution. *Springer Science and Business Media LLC*, 55–83.

Vergeer, M., Mesik, J., Baek, Y., Wilmerding, K., & Engel, S. A. (2018). Orientation-selective contrast adaptation measured with SSVEP. *Journal of Vision*, *18*, 2.

Walker, P. (1975). Stochastic properties of binocular rivalry alternations. *Perception & Psychophysics*, *18*, 467–473.

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*, 242–268.

Weilnhammer, V., Stuke, H., Hesselmann, G., Sterzer, P., & Schmack, K. (2017). A predictive coding account of bistable perception - a model-based fMRI study. *PLoS Computational Biology*, *13*, e1005536.

Wilke, M., Mueller, K.-M., & Leopold, D. A. (2009). Neural activity in the visual thalamus reflects perceptual suppression. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 9465–9470.

Wilson, H. R. (2003). Computational evidence for a rivalry hierarchy in vision. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 14499–14503.

Wunderlich, K., Schneider, K. A., & Kastner, S. (2005). Neural correlates of binocular rivalry in the human lateral geniculate nucleus. *Nature Neuroscience*, *8*, 1595–1602.

Zhang, R., Engel, S. A., & Kay, K. (2017). Binocular Rivalry: A Window into Cortical Competition and Suppression. *Journal of the Indian Institute of Science*, *97*, 477–485.

Zhou, G., Zhang, L., Liu, J., Yang, J., & Qu, Z. (2010). Specificity of face processing without awareness. *Consciousness and Cognition*, *19*, 408–412.

3.

Does Expectation Modulate Early Sensory Processing?

Predictive processing accounts suggest that prediction is an intrinsic component of all levels of sensory processing, shaping sensory representations from the earliest stages by suppressing predictable input and eliciting prediction error responses where mismatches occur. In this framework, sensory predictions are carried in cascading feedback to ‘expectation units’ at each level of analysis to meet sensory signals as they arrive, but only the prediction error arising from these comparisons is propagated forwards to the next stage of processing by ‘error units’. In contrast, traditional accounts of the computational architecture of visual cortex typically characterise feedback contributions as weak, secondary influences that serve to modify a primarily feedforward processing flow (Crick & Koch, 1998; Koch & Poggio, 1999). From this perspective, backwards connections primarily support the allocation of attentional resources (Saalman et al., 2007) and contribute to the stability of neural networks (Douglas et al., 1995) within a protected stage of early sensory processing that is free from cognitive influences (Pylyshyn, 1999). Despite these differences, traditional and predictive processing frameworks have proven deceptively difficult to dissociate empirically (Walsh et al., 2020). The present study leveraged the feature-selectivity and high temporal resolution of EEG to target a divergence in the proposed neurophysiological implementation of these processing schemes and their immediate response to predictable stimulus sequences. Specifically, predictive processing claims that sensory responses to expected stimuli should be attenuated, a phenomenon termed 'expectation suppression'. Conversely, when expectations are

not accurate, this account predicts enhanced stimulus-evoked activity, referred to as a 'surprise response', representing the swell of prediction error signalling.

Neurophysiological Evidence of Expectation Influencing Sensory Processing

A differential event-related potential (ERP) component called the mismatch negativity (MMN; Naatanen et al., 1978) is often pointed to as evidence of prediction in sensory processing (Stefanics et al., 2014; Winkler & Czigler, 2012). The majority of MMN research focuses on responses to unexpected stimuli in auditory sequences (den Ouden et al., 2012), but there is also a visual equivalent of the MMN (Stefanics et al., 2014). The visual MMN (vMMN), which exhibits a posterior scalp distribution and peaks approximately 150-400 ms after the onset of the unexpected stimulus (Kimura, 2012), has been shown to respond to the violations of both simple and abstract regularities across a wide variety of stimulus features (Stefanics et al., 2014). This complex sensitivity has led some to frame the (v)MMN as a prime candidate for an electrophysiological marker of cortical prediction error (Friston, 2005; Garrido et al., 2009; Stefanics et al., 2014, 2018; Winkler & Czigler, 2012). However, it is difficult to discern the role of expectation in many of its reported dynamics because the vMMN, like the MMN, has been most extensively studied with oddball paradigms, where the standard sequence becomes expected through repetition. When stimuli are repeatedly presented, adaptation attenuates stimulus-evoked responses (Grill-Spector et al., 2006) and the fresh afferent activity associated with a deviant stimulus may resemble a prediction error response without any influence from top-down sensory predictions (May & Tiitinen, 2010). Taken together, the lack of consensus about what the vMMN is and what mechanisms best account for its presence undermine its credibility as the electrophysiological signature of prediction error.

Traditional approaches typically assume early sensory processing exhibits a high degree of fidelity to the physical stimulus with little or no 'cognitive penetration' (Pylyshyn, 1999), but predictive processing models claim that prediction influences sensory processing from the very earliest stages (Clark, 2013; Friston, 2010). The vMMN is thought to originate in extrastriate cortex (Czigler et al., 2004; Kimura et al., 2010, 2011; Kimura, 2012; Yucel et al., 2007) and the overwhelming majority of studies discussed in Chapter 1 investigated advanced processing stages in the ventral visual stream. The few studies that have investigated expectation effects in primary visual cortex have struggled to provide definitive evidence of prediction influencing the earliest stages of processing, either because the complexity of the stimuli makes the interpretation of retinotopic V1 responses difficult (Pajani et al., 2017; Richter & de Lange, 2019; Richter et al., 2018; Utzerath et al., 2017; Walsh & McGovern, 2018), because of the potential confounding effects of repetition suppression (see Feuerriegel et al., 2021), or because of some limitations of fMRI recordings (discussed further below).

To our knowledge only one study investigated the role of expectation in sensory processing by directly recording activity in primary visual cortex. Solomon et al. (2021) presented monkeys with standardised sequences of gratings presented as a continuous stream, where the pattern was occasionally disrupted by a deviant grating, and recorded neuronal responses in V1 and V4. They searched for evidence of an expectation

violation response across several variants of the task, defining the deviant grating by changes in orientation, stimulus size, contrast, or stimulus duration. In each case, there was no evidence of a surprise response in either V1 or V4 and only a small proportion of neurons showed any modulation according to the sequence regularities. They followed up the monkey neurophysiology with a complementary EEG study in humans. In the first session, participants were asked to count the number of gratings in a 3 minute block, directing their attention to the stimulus, without making its orientation task-relevant. Again the authors found no significant difference between the occipital ERPs evoked by standard and deviant stimuli. In a second session, participants were informed about the standard sequences and told to actively detect violations of this pattern. In this case, there was a significant difference in the responses to standard and deviant stimuli, beginning ~250 ms after stimulus onset. A current source density analysis suggested an early occipital component of this response was followed by a later parietal component, providing some evidence for an error response in sensory cortex during direct task engagement with the stimulus.

Expectation and Attention

In the predictive processing framework, attention is conceptualised as optimising the influence of prediction errors on the perceptual hypothesis by adjusting their precision-weighting (see Chapter 1; Feldman & Friston, 2010; Friston, 2009). Importantly, it has been suggested that precision-weighting is implemented by adjusting the synaptic gain on error units, so attentional states will influence the manifestation of expectation suppression or prediction error responses in neural recordings (Brown & Friston, 2012; Feldman & Friston, 2010). Although many studies of expectation implicitly manipulate attention (Summerfield & Egner, 2009), most of the research on expectation has not systematically manipulated attention, commonly just requiring passive fixation or monitoring the stimulus for rare targets. The research that has attempted to investigate this relationship has provided mixed results, with some reporting that expectation effects are preserved when the stimulus is unattended (den Ouden et al., 2009), others finding attention enhances the difference between responses to expected and unexpected stimuli (Jiang et al., 2013; Smout et al., 2019), and others observing a complete reversal of the effect of expectation in attended/unattended conditions (Kok et al., 2012a).

Recently, Richter and de Lange (2019) reported that BOLD responses to expected compared to unexpected stimuli were attenuated across the ventral visual stream, but this effect was only present when the objects were directly attended, there was no expectation effect in the unattended condition (also reported by Larsson & Smith, 2012). The authors suggest that previous studies that have shown expectation effects for unattended stimuli (den Ouden et al., 2009; Kok et al., 2012a) may not have sufficiently distracted attention from the stimulus, since it was the only object on the screen in these studies, while Richter and de Lange's task used distinct stimuli at fixation that could compete for attention. St. John-Saaltink et al. (2015) also reported that expectation effects are abolished depending on the task performed by the participants. However, they found the exact opposite of Richter and de Lange, observing suppression of BOLD responses to expected grating

stimuli when participants performed a task at fixation, but not when they performed a task that required them to attend to the grating stimulus itself.

Finally, it is worth noting that predictive processing is not the only framework that can accommodate a relationship between expectation and attention. The detection of unexpected stimuli has obvious adaptive applications in signalling novelty and the salience of surprising events may serve to inform the distribution of attentional resources to identify and respond to those events. Alink and Blank (2021) propose an attention-based account of expectation effects, suggesting that the allocation of attention to surprising stimuli may enhance the gain of the stimulus-evoked activity, generating a 'surprise response'. Conversely, stimuli that are entirely in line with expectations require a lesser investment of attentional resources for processing, which might explain the relative suppression of responses to expected stimuli.

The Present Study

In this study, the contention that expectation shapes the very earliest stages of stimulus encoding was tested by measuring the responses of an established electrophysiological signature of sensory processing, the steady-state visual evoked potential (SSVEP; reviewed by Norcia et al., 2015), to expected and unexpected sequences of contrast changes in a checkerboard stimulus. The SSVEP is thought to originate in primary visual cortex when it is evoked by simple grating stimuli (Di Russo et al., 2007; Lauritzen et al., 2010; Vanegas et al., 2013) and because it is generated by flickering the stimulus at a particular frequency, the signal provides a direct measure of the oscillatory activity of sensory neurons responding to the stimulus. In addition to its excellent contrast sensitivity, the amplitude of SSVEP signals predict choice behaviour on tasks where participants are asked to detect or discriminate stimulus contrast changes, validating its characterisation as an index of the processing of sensory evidence (O'Connell et al., 2012; Steinemann et al., 2018). To account for the possibility that expectation does not influence early sensory processing, but can modulate activity at subsequent stages of the processing stream or in subsequent recurrent influences, ERPs were also analysed to determine if unexpected changes elicited a vMMN, which is thought to emerge at a later stage of visual processing (Kimura, 2012). Previous studies have been criticised for confounding expectation effects with adaptation effects associated with repetition of a standard stimulus (Feuerriegel et al., 2021). The paradigm controls for this issue because the standard sequence is composed of several contrast changes and the unexpected stimulus could be either high contrast or low contrast. Finally, the previous section illustrated that the effects of expectation may shift depending on attentional engagement with the stimulus, although it is not entirely clear in which direction this interaction should occur. To examine the nature of this relationship and limit speculation about the role of attention in any result, two experiments were conducted with two different versions of the task: Experiment 1 required attention at fixation and Experiment 2 required participants to directly attend to the stimulus feature being manipulated.

While the advantage of the excellent spatial resolution of fMRI recordings was illustrated in numerous studies described in Chapter 1, there are two primary weaknesses with this approach in the current context. First, predictive processing proposes that the dynamics of prediction error minimisation play out in two distinct neural populations and the response of error units and expectation units to predicted or surprising events should reflect this functional divergence. However, voxel-based analyses cannot segregate error units from expectation units and making predictions about the macroscopic response as the activity of these populations coalesce is difficult without a clear understanding of their relative prevalence and contribution to BOLD activity. Second, haemodynamic responses are famously sluggish and there is a somewhat murky relationship between BOLD activity and the neural activity it is taken to represent (Logothetis, 2008). The poor temporal resolution of these measurements does not allow for the very earliest sensory responses to be investigated in a timeframe before subsequent recurrent influences intervene.

To our knowledge, no other studies have examined the effects of expectation on sensory processing using the SSVEP, but it represents an improvement on both of these fronts. First, the phenomena of interest, expectation suppression and prediction error responses, should be expressed by error units and predictive processing theorists (e.g. Friston, 2009) have explicitly associated error units with the superficial pyramidal cells that are primarily responsible for generating EEG signals (Cohen, 2017). Since the neuronal activity captured in the SSVEP is directly driven by the stimulus oscillation, the signal should preferentially represent neurons tuned to the features of the stimulus. These are the cells that would be expected to exhibit the dynamics described in predictive processing accounts. This contrasts with the many fMRI studies that are forced to approximate the effects of expectation on populations with divergent stimulus preferences by estimating the average preference of voxels. Second, the temporal resolution of the SSVEP is also an improvement on that of BOLD responses. The Fourier transformation, needed to extract the signal at the frequency of interest, requires a window of activity, which means that the SSVEP does not have the same millisecond resolution of an ERP, but there is still improved scope for capturing the initial sensory response over the temporal resolution of BOLD activity. In this sense, the SSVEP offers a uniquely targeted window on the neural populations deemed most likely to demonstrate their role in prediction error minimisation.

Methodology

Participants

15 participants completed Experiment 1 (11 females, Range: 19-20, M = 23.1) and 19 completed Experiment 2 (5 females, Range: 18-35, M = 23.6). Participants completed either Experiment 1 or Experiment 2, none participated in both. All participants were over 18 years of age, reported normal or corrected-to-normal vision, no history of migraine, no history of epilepsy or psychiatric diagnosis, and no sensitivity to flashing light. The participants were students in Trinity College Dublin and were offered research credits as compensation for their time. Participants provided informed written consent prior to testing and all procedures were approved by the Trinity College Dublin School of Psychology Ethics Committee and were in accordance with the Declaration of Helsinki.

Apparatus

The experiments[1] were conducted in a dark, sound-attenuated testing booth. Visual stimuli were generated using Psychtoolbox and a custom Matlab experimental script and were presented on a 51 cm gamma-corrected CRT monitor (1024x768 resolution, 100 Hz frame rate, 65.2 cd/m2 mean luminance). A chin rest was used to reduce head movement and ensure that the viewing distance was 60 cm. Continuous EEG data were acquired from 128 scalp electrodes using a Biosemi ActiveTwo system. These signals were digitised at 512 Hz. Vertical eye movements were recorded using two vertical electrooculogram (VEOG) electrodes placed above and below the left eye.

Experimental Task

The experiments were designed to facilitate the measurement of sensory responses to identical stimuli under conditions where those stimuli were either expected or unexpected. This was achieved using a simple paradigm consisting of two short sequences of stimulus contrast changes, which could either conform to or violate previously-instilled expectations. The stimulus was an annulus with a square-wave checkerboard pattern, presented centrally on a grey background with the same mean luminance. The checkerboard had an inner radius of 3° visual angle and an outer radius of 8° of visual angle. A fixation point with a radius of 0.06° visual angle was presented in the central annulus to reduce eye movements.

[1] Experiment 1 and Experiment 2 were based on the same experimental task, which will be described in this section. Any differences between the experiments will be highlighted, all other details were common to both experiments.

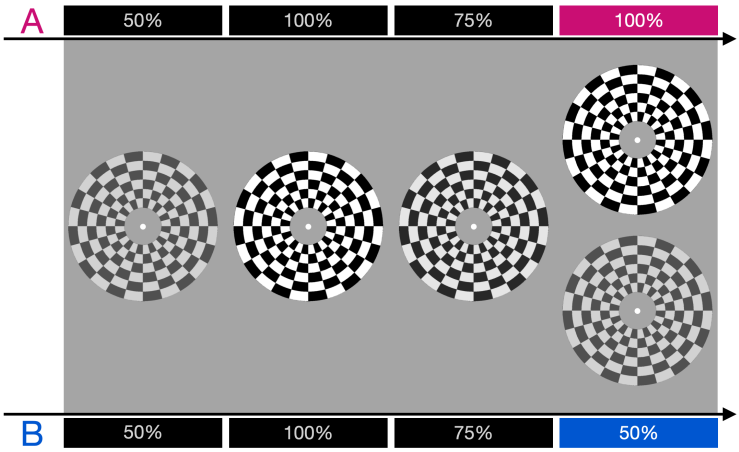


Figure 3.1. *Contrast change sequences. In both sequences, contrast progressed through four step-changes on each trial. The first three of these changes were identical for sequence A and B (50% - 100% - 75%). The final change identified the sequences as either A (100%) or B (50%). Each contrast level was presented for 600ms in Experiment 1 and 1000 ms in Experiment 2. During a practice block completed at the beginning of each testing session, all 80 trials were composed of only either sequence A or sequence B, this sequence was then labelled 'expected' and the other sequence became 'unexpected' for the remainder of the session. This arrangement was then reversed for the second testing session. The order in which each sequence was labelled expected was randomly counterbalanced across participants.*

The task used two sequences (labelled A and B in Figure 3.1) of four step-changes in stimulus contrast at regular intervals (Experiment 1: step-change every 600 ms; Experiment 2: every 1000ms). The first three of the four contrast steps were identical (50% - 100% - 75%), so the sequence was defined as A or B based on the final contrast change (50% or 100%). In each session, one of these sequences was defined as expected and the other sequence was defined as unexpected (see Procedure). Trials were separated by a 1000 ms inter-trial interval, where only the fixation point was presented on screen. To provide an index of the sensory representation of the stimulus, the checkerboard was flickered at 25 Hz by rapidly alternating segments of the checkerboard stimulus (Figure 3.2). This evoked an SSVEP, a frequency-tagged neural marker of the stimulus contrast (Norcia et al., 2015).

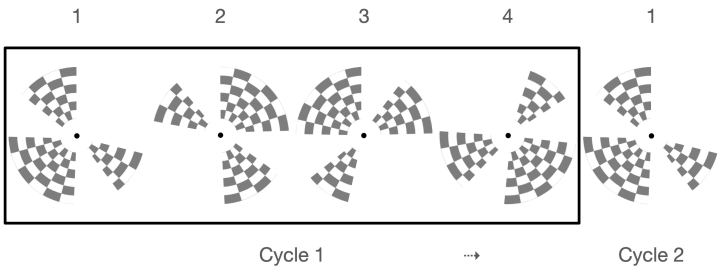


Figure 3.2. *SSVEP flicker cycle. The SSVEP was generated by flickering the stimulus at 25 Hz. The flicker was achieved by presenting a partially occluded checkerboard on each frame in a four-frame cycle composed of alternating patterns of occlusion. On each frame, 50% of the checkerboard stimulus was occluded and 50% was visible. The visible elements were divided into three segments. The portions of the checkerboard that were occluded/visible changed on each frame of the cycle.*

Electrophysiological recordings of neural responses to sensory stimuli can be affected by coincident motor activity, attentional shifts, and blinks associated with physical responses to the task. However, in the absence of an experimental task, it is difficult to ensure the participant consistently attends to the desired element of the stimulus and remains engaged throughout the session. This problem was addressed by creating a target detection task, designed to maintain attentional engagement, but not interfere with the delicate measurement of sensory responses. The literature suggests that attentional engagement may also influence any sensory modulations associated with stimulus expectations (e.g. Richter & de Lange, 2019). To account for this, the target detection task was used to direct participants' attention towards or away from the frequency-tagged stimulus (see Figure 3.3). In Experiment 1, participants were told to monitor the fixation point for a transient (400 ms) inflation, where the size of the fixation point briefly doubled. Experiment 2 involved performing a contrast-change detection task, requiring the participant to directly attend to the checkerboard stimulus itself. In this experiment, they were asked to monitor the stimulus for a sudden and distinctive fluctuation in the stimulus contrast, while maintaining fixation on the fixation point. This target consisted of a 400 ms period in which the stimulus contrast instantaneously decreased to 10% before ramping up to 90% contrast. This target was easily distinguished from the discrete step-changes characterising the stimulus contrast sequences. Both the fixation point and stimulus targets could occur at any point during the contrast change sequence. In both experiments, the participant was instructed to respond to the target as soon as they detected it. At the end of each trial, auditory feedback indicated correct responses (high-pitched beep), false alarms, or missed targets (both low-pitched beeps). If it was not a target trial and there was no response, there was no feedback. Any response recorded more than 2000 ms after target onset was labelled a miss. Critically, targets requiring a response occurred on only 10% of trials and were entirely independent of the expectational manipulation. These target trials were excluded from the subsequent analysis of neural signals, leaving 90% of trials where the participant was actively attending to the desired element of the stimulus and no response was required.

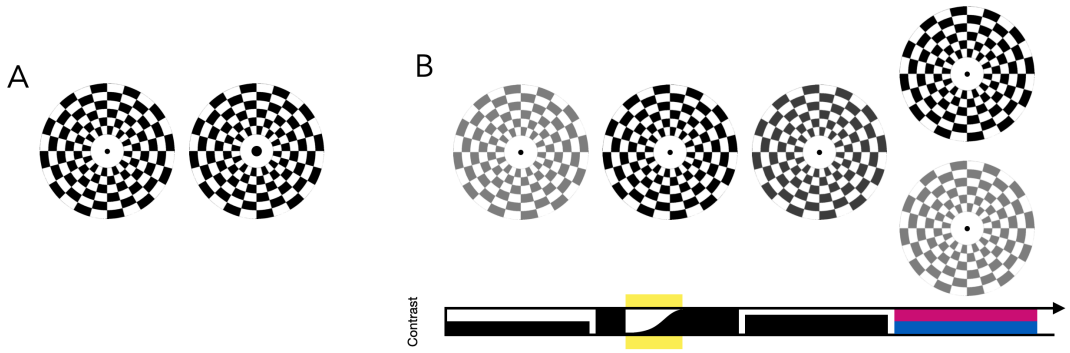


Figure 3.3. *Target stimuli. Both experiments incorporated detection tasks to direct the participant's attention towards or away from the grating stimulus. A) In Experiment 1, the participant was asked to monitor the fixation point for a sudden increase in its size. B) In Experiment 2, the participant was instructed to maintain fixation on the fixation point, but to attend to the grating stimulus for a transient fluctuation in contrast. This target, highlighted in yellow, interrupted the regular step-change sequence with a sudden decrease in contrast to 10%, followed by the contrast ramping up to 90% before returning to the pre-target contrast level. In both experiments, participants were asked to respond with a button press as soon as they detected the target.*

Procedure

In both experiments, participants were asked to complete a simple target detection task on discrete trials, responding to a sudden change in the size of the fixation point (Experiment 1) or a transient fluctuation in the contrast of the grating stimulus (Experiment 2). Participants responded using a mouse held in the palm of their hands with their thumbs resting on each mouse button. For all sessions, the participant was comfortably seated in the testing booth and placed their chin on a headrest in front of the monitor. The experimenter demonstrated how the mouse should be held, cupped in the palms with arms outstretched on the table in front of them. Participants were instructed to maintain fixation on the fixation point throughout all tasks.

Practice/Expectation Setting

To instil an expectation, participants completed a practice block of 80 trials at the beginning of both testing sessions. This block required participants to perform the same target detection task as the main experiment, but the block was entirely composed of only either sequence A or B. This sequence became the 'expected' sequence for that session and the other, unseen, sequence was labelled the 'unexpected' sequence. Importantly, the 75% contrast stimulus only occurred in the penultimate position in the sequence, so it exclusively preceded the expected contrast stimulus during training. The participant was told that they would perform a practice block to (re)familiarise them with the target detection task. They were told to watch for either the fixation (Experiment 1) or contrast (Experiment 2) target, to respond as soon as they detected the target, and that the response deadline was 2 seconds. Participants were told that these targets were relatively rare, so there would be many trials where no target occurred and they would need to remain vigilant throughout the task to spot them. Auditory feedback was also provided during the practice block. The participant was told that a high-pitched sound meant that they had correctly identified a target, while a low-pitched sound meant that they had either incorrectly pressed the button on a non-target trial (false alarm) or they had failed to press the button on a target trial (miss). 90% of trials in this block were non-target presentations of the expected sequence, the remaining 10% of trials contained a target.

Main Task

In both Experiment 1 and Experiment 2, participants completed two testing sessions on different days, each composed of eight blocks of 80 trials (not including the practice block). The only difference between these two sessions was the labelling of the stimulus sequences. The sequence labelled expected in the first session became the unexpected sequence for the second session and vice versa. The pattern chosen as the expected stimulus in the first session was randomly counterbalanced across participants. During the main task, the expected pattern of contrast changes was presented on 75% of trials, the unexpected pattern occurred on 15% of trials, and the remaining 10% of trials were target trials. The order of these trials was randomised within each

block. On target trials, if the participant had not yet responded when the stimulus sequence had ended, the trial would continue until the response deadline had expired. This period was indistinguishable for the participant from the inter-trial interval as only the fixation point was presented. The deadline was 2000 ms after the target had occurred. Targets could occur during either expected or unexpected stimulus sequences. Participants were encouraged to take a break after each block and could initiate the next block when they were ready.

Data Analysis

Reaction time and accuracy data were analysed using a custom Matlab script. The EEG analysis is detailed in the following sections. The statistical analyses primarily relied on mixed effects modelling. The mixed effect analysis was chosen to exploit the number of trials collected for each participant. Any expectational modulations of EEG signals were likely to be subtle effects, so it was believed the enhanced statistical power of this trial-by-trial analysis would be valuable. The procedure for assessing fixed effects in a mixed effects analysis is not standardised and there is some debate about the best approach in different types of datasets and experimental designs (Baayen et al., 2008). However, it has been argued that for small samples like this, the optimal approach is the estimation of degrees of freedom using a Satterthwaite approximation[2] to compute an F-statistic (Kuznetsova et al., 2017; Luke, 2017). This method was adopted for all mixed effects analyses. A random intercept was included in all mixed effects analyses to account for the repeated-measures design and control for inter-subject variability.

EEG Preprocessing

The EEG data were analysed with custom scripts in Matlab using the EEGLAB toolbox (Delorme & Makeig, 2004). The data were detrended and low-pass filtered below 40 Hz. Channels identified as uniquely noisy were recorded for each participant and each testing session using a custom channel variance analysis. These channels were interpolated using spherical splines. The data were then re-referenced offline using the average reference and segmented into stimulus-locked epochs. Reflecting the different stimulus sequence durations; in Experiment 1 the epoch extended from -1000 ms to 2800 ms relative to the onset of the stimulus sequence; in Experiment 2, the epoch was -1000 ms to 4400 ms. The epoched EEG and VEOG data were baseline corrected relative to the 300 ms interval before evidence onset. All target trials and all trials where there was a response were removed before analysis.

[2] This is in comparison to other commonly used methods for testing fixed effects, including the likelihood ratio test for competing models and assessing the significance of the individual predictors in the model using Wald t-values.

Typically, artifact rejection is carried out on EEG data on a trial-to-trial basis to minimise the influence of non-neural contributions to the recordings (e.g. muscle activity associated with blinks). However, it was important to retain as many of the unexpected contrast changes as possible because, by their nature, there were relatively few of these trials for each participant. This was achieved by segmenting the trials such that separate epochs were extracted for each step in the contrast sequence and artifact rejection was conducted independently on each stimulus element. By segmenting the trial this way, sequence segments that contained artifacts could be rejected without rejecting the entire trial, isolating those rare unexpected final contrast changes from any artifacts occurring earlier in the trial. For this reason, the epoched data were segmented into the four constituent elements of each contrast sequence with a 400 ms buffer on either side of each segment window. For example, in Experiment 1, the second element of the sequence was a 100% contrast stimulus presented from 600 ms to 1200 ms after stimulus onset. In the EEG analysis, this element was isolated as the window 200 ms to 1600 ms relative to stimulus onset. The 400 ms buffer was included on either side of the element window to accommodate the time-frequency analyses used to extract the SSVEP signal (see below). If the difference in activity between the VEOG channels exceeded an absolute value of 250 μV (likely blinks) or if the voltage recorded by any scalp electrode exceeded 100 μV at any time during the segment, that segment was excluded. To compensate for the effects of volume conduction across the scalp, each segment was subjected to a Current Source Density (CSD) transformation (Kayser & Tenke, 2015). This technique is used to minimise the spatial overlap between functionally distinct EEG components (Kelly & O'Connell, 2013).

SSVEP

The Short Time Fourier Transform (STFT) procedure was used to decompose the recording of neural activity into its time-frequency components. The STFT window was 400 ms, which was chosen to accommodate ten cycles of the 25 Hz SSVEP signal. The window was moved along the length of the epoch in steps of 50 ms, providing an estimate of the power of neural activity at the SSVEP frequency. This SSVEP amplitude was then normalised by subtracting the amplitude of activity in the neighbouring frequency bins to isolate the stimulus-driven signal.

Two electrodes for each participant and each session were individually selected by ranking a pool of occipital electrodes (see Chapter Appendix) according to three selection criteria. Candidate electrodes were first ranked based on the average amplitude of 25 Hz activity during stimulus presentation. Next, electrodes were ranked according to their discrimination of the third contrast stimulus (75% contrast) from the first contrast stimulus (50% contrast), and their discrimination of the third contrast stimulus from the second contrast stimulus (100% contrast). Since these elements of the stimulus sequence were common to both sessions, this process ensured that the activity at the selected electrodes closely reflected the neural representation of the contrast of the stimulus while remaining independent of the representation of the final stimulus. These three ranks were summed and the top two electrodes were selected for each participant and session.

The SSVEP signal extraction was conducted using two methods to facilitate complementary analyses. The first method involved averaging the ERPs within conditions and then running the STFT on the averaged data. This technique helps to minimise the contamination of the SSVEP signal by incidental activity[3] (e.g. motor-related potentials). A repeated measures ANOVA was then used to analyse this averaged data. The second approach involved performing the STFT on a trial-by-trial basis, making it possible to perform a mixed effect analysis, which offers superior statistical power than the ANOVA and is robust to differences in trial counts across conditions. Performing both analyses offered improved fidelity of the SSVEP, while also ensuring that any result was not due to issues with statistical power. In addition, it was important to assess whether the presence or absence of an expectational effect on the SSVEP could be considered evidence for or against the hypothesis that expectations influence basic sensory processing. To this end, a Bayesian one-way repeated measures ANOVA was used to provide a Bayes factor for the difference between expected and unexpected SSVEPs. The Bayes factor signifies the extent to which the observed data represent evidence for the null or alternative hypothesis.

Visual Mismatch Negativity

The vMMN was computed by comparing ERPs in response to the final contrast change for the expected and unexpected sequences within each recording session (e.g. subtracting the response to an expected 100% contrast stimulus from the response to an unexpected 50% contrast stimulus). The average ERP evoked by the expected final element was subtracted from the average ERP in response to the unexpected contrast change, producing a difference waveform containing the vMMN component.

Many studies using the vMMN include a pool of parieto-occipital electrodes and several latencies in their analyses as factors (e.g. Stefanics et al., 2011, 2018). However, the latency and specific topography of the vMMN was not pertinent to the experimental questions in this study, rather, its presence or absence was the primary focus. In addition, selecting a pool of electrodes to include as factors in a statistical analysis is less elegant when using a 128-electrode system (many of the aforementioned studies identify their occipital electrode pool according to the 10-20 system). Instead, four electrodes were selected for each participant and session based on an algorithmic analysis of the difference waveform.

The algorithm was designed to identify the characteristic negative peak of the vMMN by ranking the pool of occipital electrodes according to two aspects of the difference waveform: negative amplitudes in the window of interest (Rank 1) and an analysis of peaks (Rank 2). Before ranking, the difference waveform was baselined by the mean amplitude of the difference waveform in that channel across the full duration of the final stimulus segment. The baselining was intended to prevent channels with constant negative amplitudes from achieving

[3] The topography of the difference in 25 Hz activity between method 1 and method 2 (described above) showed that the activity eliminated using the first approach was primarily from central and frontal channels, supporting the argument that this approach reduces non-sensory contributions to the extracted SSVEP signal.

high rankings. First, candidate channels were ranked by the magnitude of negative activity in the difference waveform 100:300 ms after onset of the final stimulus segment (Rank 1). This was designed to identify channels containing isolated negativities in the most prominent period of the vMMN (Stefanics et al., 2014). Rank 2 was composed of two measures of the peaks in the difference waveform. Firstly, all negative peaks occurring 150 ms after stimulus onset in the difference waveform of each candidate channel were identified and their magnitude relative to the mean absolute magnitude of activity in the 100:300 ms window was recorded. Secondly, the mean absolute magnitude of activity in the difference waveform from 50 ms after the peak to 300 ms post stimulus onset was also recorded. Rank 2 was calculated by subtracting the mean post-peak amplitude from the peak amplitude and ranking the highest scoring peak for each candidate channel. Again, this was intended to isolate channels containing the discrete negativity of the vMMN followed by a return to baseline in the difference waveform from channels with regular oscillations or lots of noise in their difference waveform. Finally the Rank 1 and Rank 2 were summed. The top four electrodes were selected for each participant and session.

The vMMN was analysed using a mixed effects analysis of the ERPs in the 100:300 ms window after the onset of the final stimulus segment in each condition, baselined by the activity from -50:0 ms. The vMMN is a negative difference waveform, created by subtracting the ERP in response to the expected stimulus from the ERP in response to the unexpected stimulus. The negativity arises from pronounced negative deflections of the unexpected ERPs. If unexpected ERPs had significantly lower amplitudes than expected ERPs in the window characteristic of the vMMN, the 'expectation' factor in the mixed effects analysis would be significant and one could infer that there was a vMMN. The single-trial analysis approach used in these statistical analyses, aimed to minimise the influence of trial count differences across conditions and optimise statistical power.

Results

Behaviour

Reaction time and accuracy on the target detection task were used to confirm that participants were engaged with the stimulus throughout the testing sessions and the attentional manipulation had been effective. If participants quickly and accurately identified the rare targets associated with a particular element of the stimulus, they were likely to have been paying close attention to that part of the stimulus throughout that testing session. Participants reliably detected the targets in both experiments (Experiment 1 M: 99.01%, SE: 0.0045; Experiment 2 M: 89.32%, SE: 0.0253). Additionally, participants' reaction times (Experiment 1 M: 512 ms, SE: 26.5 ms; Experiment 2 M: 849.6 ms, SE: 28 ms) indicated that they had been carefully monitoring the stimulus allowing them to respond rapidly.

SSVEP

For all SSVEP analyses, the SSVEP was baselined by the average activity in the first three segments of the stimulus sequence, which were identical in all conditions. This baselined SSVEP response to the four contrast elements is shown for Experiments 1 and 2 in Figure 3.4. All analyses were based on the amplitude of the SSVEP measured in the window 50:300 ms (Experiment 1) or 50:500 ms (Experiment 2) after the onset of the final stimulus segment.

First, a repeated measures ANOVA was conducted on the averaged SSVEP amplitude for each condition. In Experiment 1, the SSVEP amplitude was significantly greater when the contrast of the final segment was 100% compared to 50% ($F(1,14) = 7.067$, $p = 0.019$), but there was no main effect of expectation ($F(1,14) = 0.343$, $p = 0.568$) and no interaction ($F(1,14) = 0.001$, $p = 0.982$). In Experiment 2, the SSVEP amplitude was also significantly greater when the contrast of the final stimulus segment was 100% ($F(1,18) = 7.151$, $p = 0.015$), and again there was no main effect of expectation ($F(1,18) = 0.174$, $p = 0.682$) and no interaction ($F(1,18) = 0.325$, $p = 0.576$). In each case, the non-significant numerical trend was towards larger responses for expected stimuli.

This comparison was repeated with a mixed effects approach to analyse the SSVEP on a trial-by-trial basis. In Experiment 1, there was a significant main effect of contrast level ($F(1,13224.361) = 331.952$, $p < 0.001$), but no main effect of expectation ($F(1,13224.289) = 2.819$, $p = 0.093$) and no interaction ($F(1,13229.098) = 0.295$, $p = 0.587$). In Experiment 2, there was also a significant main effect of contrast level ($F(1,13933.504) = 98.512$, $p < 0.001$), no main effect of expectation ($F(1,13933.652) = 0.547$, $p = 0.46$), and no interaction ($F(1,13949.352) = 0.002$, $p = 0.96$). In both experiments, the SSVEP amplitude was significantly greater when the final stimulus segment was presented at 100% contrast than when it was shown at 50% contrast.

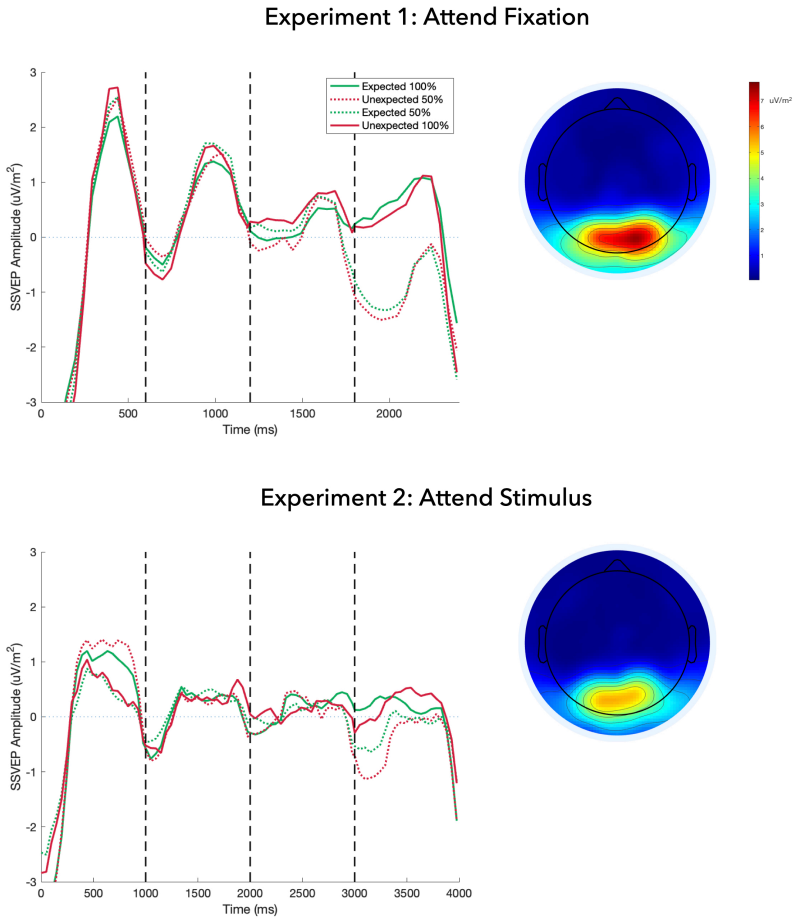


Figure 3.4. *SSVEP responses in Experiment 1 and 2. The data shown here were used in the mixed effects analysis. The STFT was run on each trial separately. The SSVEP was then normalised by subtracting neighbouring frequencies, averaged across trials, and baselined by the mean amplitude of the signal in the first three contrast segments. The dashed markers indicate the transition from each stimulus segment to the next. While the signal does not accurately reflect the contrast levels being presented (50% - 100% - 75% - 100/50%), there are clear responses to each of the contrast changes. The topographies, which are plotted on the same scale, show the mean SSVEP activity across the entire stimulus sequence, indicating that the signal originates over occipital electrodes. The waveforms of the data used for the ANOVA analysis, where the ERPs were averaged before the STFT, were very similar.*

A Bayesian one-way repeated measures ANOVA was used to assess whether the data could be considered evidence that expectation did not influence the sensory representation of the stimulus encoded in the SSVEP. The data were averaged across contrast levels, so the analysis simply compared the mean SSVEP amplitude for expected and unexpected stimuli. This produced a Bayes factor of 0.32 for Experiment 1 and 0.48 for Experiment 2, both are considered anecdotal evidence for the null hypothesis.

In brief, this indicates that the SSVEP was sufficiently sensitive to discriminate the contrast level of the final sequence segment, but there was no difference in the sensory representation of that stimulus depending on whether the contrast change was expected or unexpected.

vMMN

Figure 3.5 shows the difference waveform created by subtracting the ERP in response to the expected final stimulus segment from the ERP in response to the unexpected final stimulus segment in each session. The vMMN is clearly visible as a negative deflection in both waveforms. The presence of the vMMN was statistically confirmed by running a mixed effects analysis on the ERPs in each condition in the window 100:300 ms after the onset of the final stimulus segment. In Experiment 1, the amplitude of ERPs were significantly more negative when the contrast change was unexpected compared to expected contrast changes ($F(1,13226.067) = 41.041$, $p < 0.001$). There was no main effect of contrast ($F(1,13226.706) = 0.321$, $p = 0.571$) and no interaction ($F(1,13230.962) = 2.688$, $p = 0.101$). Similarly, in Experiment 2, the amplitude of the response to an unexpected contrast change was significantly more negative compared to the response to an expected stimulus ($F(1,13937.095) = 55.175$, $p < 0.001$). Once more, there was no main effect of contrast ($F(1,13936.238) = 1.021$, $p = 0.312$) and no interaction ($F(1,13727.865) = 1.305$, $p = 0.253$).

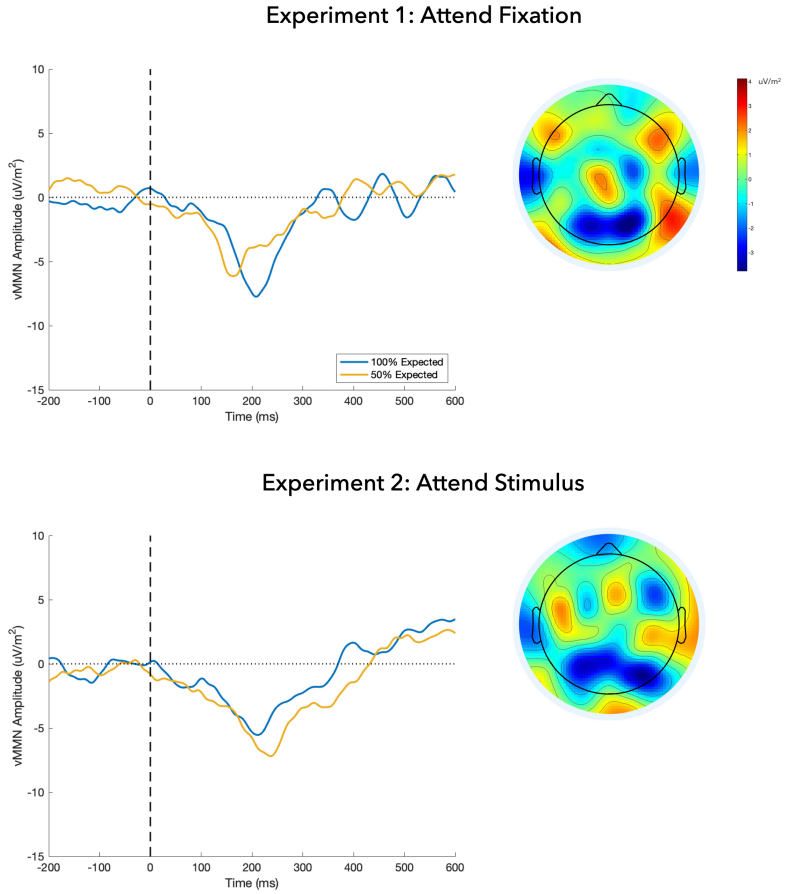


Figure 3.5. *vMMN in Experiments 1 and 2. The vMMN is a negativity in the difference waveform comparing the responses to expected and unexpected stimuli within each session. The dashed line indicates the onset of the final contrast segment. The vMMN was measured in the difference waveform from 100:300 ms. The topographies show the difference activity in a 20 ms window centred on the peak of the vMMN. It is clear that the mismatch results in a negativity over occipital electrodes. The topographies share the same scale.*

The significantly more negative ERP amplitudes for unexpected trials means that the difference waveform contains the negative deflection that can be seen in Figure 3.5. These results provide evidence that the statistics of the task were encoded and that the violation of expectations was sufficiently salient to evoke a vMMN from the visual system.

Discussion

To test the hypothesis that sensory expectation influences the earliest stages of sensory processing, responses to expected and unexpected patterns of contrast change were indexed with a well-established electrophysiological marker of early sensory encoding driven by stimulus oscillations (Norcia et al., 2015; O'Connell et al., 2012). The SSVEP exhibited a clear occipital source and a marked response to each contrast change. However, the amplitude of the SSVEP response to a contrast change did not differ depending on whether that change was expected or unexpected, providing no evidence that predictive mechanisms shaped early sensory encoding according to the sensory regularities of the stimulus sequence. A Bayesian analysis categorised this result as anecdotal evidence for the null hypothesis.

Some previous studies that have found mixed evidence of expectation effects in V1 have used complex object stimuli (e.g. Richter et al., 2018), but the stimulus used in the present task was specifically chosen to target simple contrast-sensitive cells in primary visual cortex. It is difficult to explain why numerous fMRI studies have reported expectation effects in V1 using similar stimuli to the current study (Aitken et al., 2020; den Ouden et al., 2009; Kok et al., 2012a, 2012b, 2013, 2016; St. John Saaltink et al., 2015) but the answer is likely to be due to differences between the neural activity indexed by BOLD signals and the SSVEP (outlined in the Introduction). Instead, the SSVEP is more closely related to local field potentials (Kim et al., 2007). Consistent with the results reported here, the only monkey study to have investigated the role of prediction in primary visual cortex found no evidence of expectation effects in either local field potentials or single-unit recordings in response to simple grating stimuli that violated regularities across a variety of stimulus dimensions, including contrast (Solomon et al., 2021).

It might be suggested that the present result arose because the sequential regularities had not been sufficiently salient for the participants to encode the contrast pattern in training or while they were performing the task. This would be surprising because research on statistical learning suggests that humans are highly sensitive to stimulus regularities and rapidly develop behavioural sensitivity to implicitly learned predictable patterns (Fiser & Aslin, 2002; Turk-Browne et al., 2009, 2010). Here, we were able to confirm that the stimulus sequences were registered in the brain with the detection of a clear vMMN in response to sequence violations (Figure 3.5). This provides strong evidence that the regularity was encoded at some level of the visual system and that the violation of that pattern influenced the sensory response to the stimulus. While the SSVEP is thought to originate in primary visual cortex in response to simple contrast stimuli like the checkerboard stimulus (Di Russo et al., 2007; Lauritzen et al., 2010; Vanegas et al., 2013), the vMMN is believed to originate at a later, extrastriate stage of visual processing (Kimura, 2012). Taken together, the results of the present study suggest that expectation does shape sensory processing, but does not appear to influence the earliest sensory representations.

Previous research has demonstrated that there is a complicated relationship between the influence of expectation and attention on sensory processing, which may be interactive under certain circumstances (Kok et al., 2012a). To examine the contribution of attention to these early dynamics and to control for its influence on any expectation effect, or absence thereof, the study included two experiments. One experiment required participants to attend to the fixation point, making the encompassing checkerboard stimulus task-irrelevant; a second experiment required careful monitoring of the contrast level of the checkerboard stimulus itself, making the (un)expected stimulus feature task-relevant. The attentional manipulation did not change any of the results; in both conditions the SSVEP significantly differed according to the final contrast level, but showed no expectation effect, and the unexpected contrast change evoked a vMMN.

Surprising events may attract attention exogenously (Alink & Blank, 2021), so participants' attention may have been drawn to the deviant stimulus during the fixation task. This seems unlikely because there was no difference in the SSVEP response between expected and unexpected contrast changes in this study and if there was a discrepancy in terms of attention paid to deviant and standard stimulus sequences, this would be expected to manifest as a difference in the sensory signals (Di Russo & Spinelli, 2002; Kim et al., 2007). However, in order for attention to have modulated the sensory signals in the immediate period after the deviant contrast stimulus onset, the visual system would have had to detect that this was a deviant stimulus based on an expectation and this would have been evidence that expectation shapes early sensory processing. There is also some evidence that regularities themselves attract attention (Zhao et al., 2013). However, the clear difference in the SSVEP between the conditions suggests that the attentional engagement with the stimulus was effectively manipulated by the task despite the presence of a regularity in both attention conditions.

Despite the theoretical advantage of the SSVEP outlined in the Introduction, one clear limitation in this study is that the signal failed to faithfully reflect the presented contrast levels. As can be seen in Figure 3.4, the initial response to the 50% contrast level is greater than the amplitude achieved with the subsequent 100% contrast stimulus. This inconsistency can be attributed to the abrupt onset of the contrast stimulus provoking a visual evoked potential. Smaller transient responses to the subsequent contrast changes may also explain the dips that can be seen in the SSVEP immediately after each transition, as a brief visual evoked potential would likely reduce the signal-to-noise ratio of the SSVEP. However, the responses to the subsequent stimulus transitions also fail to clearly differentiate the relative changes in contrast, particularly in Experiment 2. This was unexpected because the SSVEP is well established as a neural index of contrast-sensitive stimulus encoding (Norcia et al., 2015; O'Connell et al., 2012; Steinemann et al., 2018), but its inconsistent responses to the stimulus contrast suggests that the piecemeal stimulus modulation used to generate the SSVEP in this study may have been suboptimal for driving strongly discriminant responses. Additionally, it is quite possible that contrast adaptation is responsible for the decreasing amplitude of the SSVEP responses to the contrast changes across the sequence. Adaptation would be likely to exert a greater influence in Experiment 2, where each sequence lasted four seconds, potentially accounting for the less pronounced response to each contrast change in that experiment. While the contrast-fidelity of the SSVEP is an important caveat in interpreting these

results, there are still clear SSVEP responses to the stimulus contrast changes and these provide no indication that the changes were differentially encoded based on prior expectation.

References

Aitken, F., Menelaou, G., Warrington, O., Koolschijn, R. S., Corbin, N., Callaghan, M. F., & Kok, P. (2020). Prior expectations evoke stimulus-specific activity in the deep layers of the primary visual cortex. *PLoS Biology*, *18*, e3001023.

Alink, A., & Blank, H. (2021). Can expectation suppression be explained by reduced attention to predictable stimuli? *Neuroimage*, *231*, 117824.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Brown, H. R., & Friston, K. J. (2012). Dynamic causal modelling of precision and synaptic gain in visual perception - an EEG study. *Neuroimage*, *63*, 223–231.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204.

Cohen, M. X. (2017). Where does EEG come from and what does it mean? *Trends in Neurosciences*, *40*, 208–218.

Crick, F., & Koch, C. (1998). Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature*, *391*, 245–250.

Czigler, I., Balázs, L., & Pató, L. G. (2004). Visual change detection: event-related potentials are dependent on stimulus location in humans. *Neuroscience Letters*, *364*, 149–153.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex*, *19*, 1175–1185.

Di Russo, F, & Spinelli, D. (2002). Effects of sustained, voluntary attention on amplitude and latency of steady-state visual evoked potential: a costs and benefits analysis. *Clinical Neurophysiology*, *113*, 1771–1777.

Di Russo, Francesco, Pitzalis, S., Aprile, T., Spitoni, G., Patria, F., Stella, A., … Hillyard, S. A. (2007). Spatiotemporal analysis of the cortical sources of the steady-state visual evoked potential. *Human Brain Mapping*, *28*, 323–334.

Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A., & Suarez, H. H. (1995). Recurrent excitation in neocortical circuits. *Science*, *269*, 981–985.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215.

Feuerriegel, D., Vogels, R., & Kovács, G. (2021). Evaluating the evidence for expectation suppression in the visual system. *Neuroscience and Biobehavioral Reviews*, *126*, 368–381.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*, 458–467.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*, 815–836.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*, 293–301.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*, 127–138.

Garrido, M. I., Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2009). Dynamic causal modeling of the response to frequency deviants. *Journal of Neurophysiology*, *101*, 2620–2631.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*, 14–23.

Jiang, J., Summerfield, C., & Egner, T. (2013). Attention sharpens the distinction between expected and unexpected percepts in the visual brain. *The Journal of Neuroscience*, *33*, 18438–18447.

Kayser, J., & Tenke, C. E. (2015). On the benefits of using surface Laplacian (current source density) methodology in electrophysiology. *International Journal of Psychophysiology*, *97*, 171–173.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *The Journal of Neuroscience*, *33*, 19434–19441.

Kimura, M., Ohira, H., & Schröger, E. (2010). Localizing sensory and cognitive systems for pre-attentive visual deviance detection: an sLORETA analysis of the data of Kimura et al. (2009). *Neuroscience Letters, 485*, 198 – 203.

Kimura, M., Kondo, H., Ohira, H., & Schröger, E. (2011). Unintentional temporal-context based prediction of emotional faces: an electrophysiological study. *Cerebral Cortex, 22(8),* 1774-1785.

Kimura, M. (2012). Visual mismatch negativity and unintentional temporal-context-based prediction in vision. *International Journal of Psychophysiology*, *83*, 144–155.

Kim, Y. J., Grabowecky, M., Paller, K. A., Muthu, K., & Suzuki, S. (2007). Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nature Neuroscience*, *10*, 117–125.

Koch, C., & Poggio, T. (1999). Predicting the visual world: silence is golden. *Nature Neuroscience*, *2*, 9–10.

Kok, P., Brouwer, G. J., van Gerven, M. A. J., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *The Journal of Neuroscience*, *33*, 16275–16284.

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012b). Less is more: expectation sharpens representations in the primary visual cortex. Neuron, 75, 265–270.

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012a). Attention reverses the effect of prediction in silencing sensory signals. Cerebral Cortex, 22, 2197–2206.

Kok, P., van Lieshout, L. L. F., & de Lange, F. P. (2016). Local expectation violations result in global activity gain in primary visual cortex. *Scientific Reports*, *6*, 37706.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmertest package: tests in linear mixed effects models. Journal of Statistical Software, 82.

Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cerebral Cortex*, *22*, 567–576.

Lauritzen, T. Z., Ales, J. M., & Wade, A. R. (2010). The effects of visuospatial attention measured across visual cortex using source-imaged, steady-state EEG. *Journal of Vision*, 10.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*, 869–878.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494–1502.

May, P. J. C., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*, 66–122.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta psychologica*, 42(4), 313-329.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*, 4.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*, 1729–1735.

Pajani, A., Kouider, S., Roux, P., & de Gardelle, V. (2017). Unsuppressible Repetition Suppression and exemplar-specific Expectation Suppression in the Fusiform Face Area. *Scientific Reports*, *7*, 160.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*, 341–365; discussion 366.

Richter, D., & de Lange, F. P. (2019). Statistical learning attenuates visual activity only for attended stimuli. ELife, 8.

Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed Sensory Response to Predictable Object Stimuli throughout the Ventral Visual Stream. *The Journal of Neuroscience*, *38*, 7452–7461.

Saalmann, Y. B., Pigarev, I. N., & Vidyasagar, T. R. (2007). Neural mechanisms of visual attention: how top-down feedback highlights relevant locations. *Science*, *316*, 1612–1615.

Smout, C. A., Tang, M. F., Garrido, M. I., & Mattingley, J. B. (2019). Attention promotes the neural encoding of prediction errors. *PLoS Biology*, *17*, e2006812.

Solomon, S. S., Tang, H., Sussman, E., & Kohn, A. (2021). Limited evidence for sensory prediction error responses in visual cortex of macaques and humans. *Cerebral Cortex*, *31*, 3136–3152.

St. John-Saaltink, E., Utzerath, C., Kok, P., Lau, H. C., & de Lange, F. P. (2015). Expectation suppression in early visual cortex depends on task set. Plos One, 10, e0131172.

Stefanics, G., Heinzle, J., Horváth, A. A., & Stephan, K. E. (2018). Visual Mismatch and Predictive Coding: A Computational Single-Trial ERP Study. The Journal of Neuroscience, 38, 4020–4030.

Stefanics, G., Kimura, M., & Czigler, I. (2011). Visual mismatch negativity reveals automatic detection of sequential regularity violation. Frontiers in Human Neuroscience, 5, 46.

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. Frontiers in Human Neuroscience, 8, 666.

Steinemann, N. A., O'Connell, R. G., & Kelly, S. P. (2018). Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nature Communications*, *9*, 3627.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*, 403–409.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*, 1934–1945.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by

statistical learning. *The Journal of Neuroscience*, *30*, 11177–11187.

Utzerath, C., St John-Saaltink, E., Buitelaar, J., & de Lange, F. P. (2017). Repetition suppression to objects is modulated by stimulus-specific expectations. *Scientific Reports*, *7*, 8781.

Vanegas, M. I., Blangero, A., & Kelly, S. P. (2013). Exploiting individual primary visual cortex geometry to boost steady state visual evoked potentials. *Journal of Neural Engineering*, 10, 036003.

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*, 242–268.

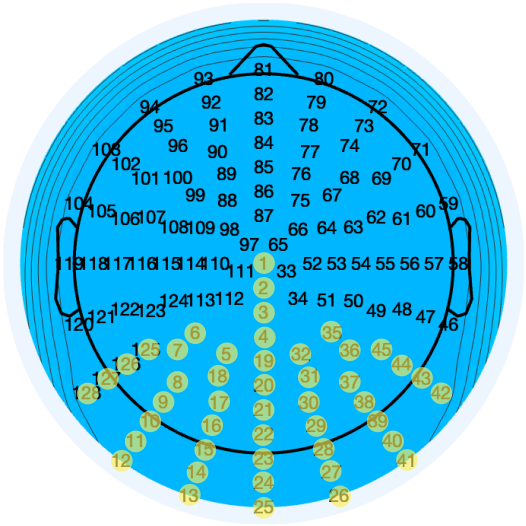
Walsh, K. S., & McGovern, D. P. (2018). Expectation suppression dampens sensory representations of predicted stimuli. *The Journal of Neuroscience*, *38*, 10592–10594.

Winkler, I., & Czigler, I. (2012). Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *International Journal of Psychophysiology*, *83*, 132–143.

Yucel, G., McCarthy, G., & Belger, A. (2007). fMRI reveals that involuntary visual deviance processing is resource limited. *Neuroimage*, *34*, 1245–1252.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, *24*, 667–677.

Appendix 3.1 - Electrode Selection



The electrodes for each participant were selected from the pool of highlighted electrodes above. The specific electrode selections for each experiment are listed in the table on the right.

Experiment 1		**Experiment 2**	
Session 1	Session 2	Session 1	Session 2
18, 8	45, 36	15, 16	15, 23
14, 15	13, 29	29, 24	29, 30
20, 22	19, 21	17, 23	25, 24
17, 18	17, 18	19, 32	21, 22
13, 14	10, 36	26, 31	23, 24
17, 18	17, 18	21, 22	25, 18
19, 22	20, 21	15, 37	31, 32
23, 22	23, 22	23, 8	5, 7
20, 21	5, 19	29, 30	24, 25
4, 19	20, 21	30, 31	29, 30
17, 18	17, 18	20, 21	16, 28
17, 18	17, 18	42, 43	39, 40
13, 23	22, 21	16, 17	23, 32
29, 28	29, 30	25, 30	28, 29
31, 30	29, 30	26, 27	12, 27
		19, 29	19, 37
		30, 31	30, 31
		24, 25	23, 24
		25, 38	21, 22

4.

Perceptual Decision-Making and The Drift Diffusion Model

Perceptual decision-making is the process by which animals convert ambiguous sensory information into useful interpretations of their environment (Gold & Shadlen, 2007). Although the rigorously constrained experimental paradigms and simplified psychophysical stimuli used to investigate perceptual decision-making in laboratory settings can appear abstract, niche, and detached from normal human experience, this is merely a reflection of the limitations of our knowledge and technology. Scientists explain complex phenomena with incremental advances from the most rudimentary models and there are no known phenomena more complex than the foundations of human cognition. However, the trade-offs made in ecological validity can obscure the underlying reality that this field of research aims to illuminate the mechanisms behind one of the most pervasive and fundamental components of conscious experience (Shadlen & Kiani, 2013). In a lab, perceptual decision-making might entail thousands of trials distinguishing subtle patterns of motion in clouds of noisy dots, but this is an abstraction of the same task faced by drivers in foggy conditions, skiers in rapid descent, and bleary-eyed PhD students searching for the snooze button. In real life, every set of traffic lights, every glimpsed face, every whiff of coffee, and every notification vibration is another perceptual decision made with such seamless efficiency that often they do not even register as conscious processes.

It is precisely because they are so finely woven into the fabric of our daily experience that, over the last 30 years, cognitive neuroscientists have doggedly pursued an understanding of the simplest perceptual decisions

(O'Connell & Kelly, 2021; Summerfield & Blangero, 2017). The enthusiasm stems from the prospect of exploiting a tractable cognitive phenomenon to construct a unified neurophysiological and computational model of a higher level mental-process; to discover a "window on cognition" (Shadlen & Kiani, 2013). Indeed, perceptual decisions are subject to many of the same constraints and influences that characterise all kinds of decisions (e.g. time pressure, value biases, prior experience, confidence). The union of well-studied psychophysical paradigms, fine-grained neurophysiological recordings, and computational models of perceptual decision-making has been a catalyst for significant advances in cognitive neuroscience (Gold & Shadlen, 2007; Hanks & Summerfield, 2017). In particular, sequential sampling models have provided a principled framework for the investigation of the dynamics underlying the decision process (Ratcliff et al., 2016). This is a diverse model family (Ratcliff & Smith, 2004), but the Drift Diffusion Model has emerged as the dominant variant (Forstmann et al., 2016; Gold & Stocker, 2017).

The Drift Diffusion Model conceptualises decision-making as a dynamic process of accumulation of noisy evidence over time (Ratcliff & McKoon, 2008; see Figure 4.1). The decision is made and a response initiated when the decision variable hits one of two decision boundaries. Since the Drift Diffusion Model only allows for two choice alternatives, evidence for one hypothesis is necessarily evidence against the alternative. Therefore,

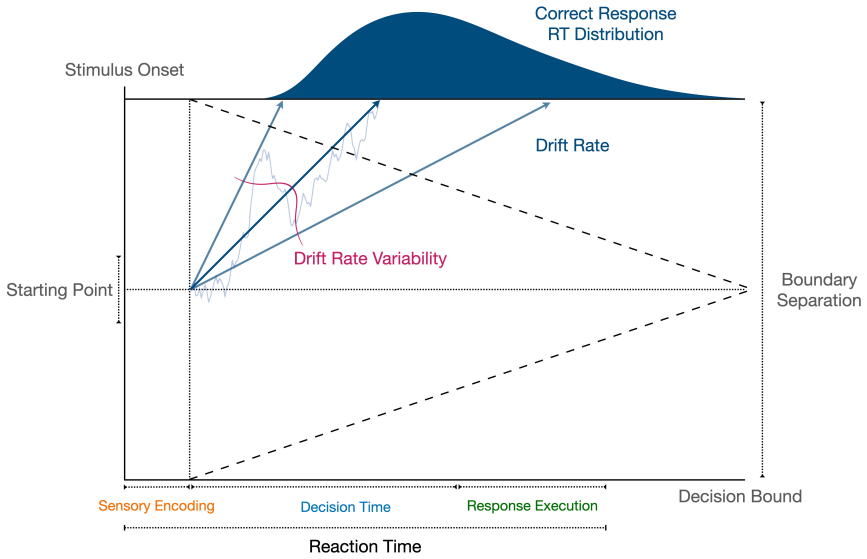


Figure 4.1. *The Drift Diffusion process for a set evidence strength. After a period of sensory encoding, evidence begins to accumulate at a rate proportionate to the signal-to-noise ratio. Noise in sensory encoding produces variability in this drift rate. Additional noise in the representation of the accumulated evidence disturbs the trajectory of the decision variable predicted by the drift rate (light blue trace illustrates the effect of this within trial variability). When the decision variable reaches one of the bounds, the decision process terminates and a response is initiated. The delay associated with motor execution is the final component of reaction time. The noise in the process results in a distribution of reaction times. Here, the upper bound is arbitrarily assigned to the correct response. Although the outer drift rates in the illustration represent the extremes of the drift rate distribution for a single evidence strength, the effect would be equivalent for varying evidence strength. Strong evidence would produce the steepest ascent and narrower drift rate variability; weaker evidence would produce the slowest ascent and greater drift rate variability. The Drift Diffusion Model has static bounds, but some have argued that bounds collapse over the course the decision to minimise the costs associated with deliberation (dashed lines). This bound collapse can be neurophysiologically implemented in a non-selective urgency signal at the motor level, pushing preparation for both responses towards the bound.*

the sampled evidence can be encoded by a single decision variable, which acts as a temporal integrator representing the relative evidence favouring one alternative over the other. Following a delay after evidence onset as the initial sensory information is processed, the decision variable embarks on its journey towards one of the two decision boundaries from a starting point. The separation of the bounds determines the quantity of evidence needed to make a decision; greater separation produces slower, but more accurate responses.

How quickly the decision variable travels to the bound is determined by the drift rate, which represents the amount of evidence accumulated per unit of time. Strong evidence produces a greater drift rate and weak evidence produces a smaller drift rate. The process of accumulating evidence will reduce the influence of physical noise in the stimulus over time (Gold & Stocker, 2017), but the neurons encoding the sensory evidence and those encoding the state of the decision process are also noisy. The sensory noise is captured in drift rate variability and the decision noise is represented by a within-trial variability parameter. Together, this noise results in fluctuations in the decision variable's trajectory. This variability in its path to the bound produces reaction time distributions, since the decision variable will cross the bound at different times on different trials, even when they have an identical evidence strength. It also produces errors because the decision variable will be sufficiently diverted by noise on some trials and driven to the incorrect bound.

Reaction time is composed of the time taken for this decision process to unfold and non-decision time. Non-decision time consists of the time taken to encode the sensory evidence and the delay associated with execution of the motor response. Typically the bound separation is fixed throughout the deliberation period, but some have argued that a time-dependent bound collapse is called for under certain conditions (Gold & Stocker, 2017). In this case, less evidence is needed to terminate the decision process as more time elapses. Collapsing bounds may represent a more economic strategy for the investment of neural resources (Cisek et al., 2009; Drugowitsch et al., 2012), they may be important for producing decisions under time pressure (Martinez-Rodriguez et al., 2020; Steinemann et al., 2018), and they have been implicated in the incorporation of priors in the decision process (Hanks et al., 2011; Kelly et al., 2021). However, the need to incorporate bound collapse in decision models is still a contested issue (Boehm et al., 2020; Hawkins et al., 2015; O'Connell & Kelly, 2021).

Neurophysiological Evidence of Evidence Accumulation

One of the defining characteristics of cognition is its operation across different time-scales. Unlike basic sensory reflexes, cognitive function is flexible even when confronted with complex, evolving evidence. In the case of perception, if this flexibility is to extend beyond the temporal constraints biology places on sensory neurons, the brain needs to maintain a dynamic representation of previous samples of sensory evidence and update this representation as new evidence arrives over time. Sequential sampling models capture this capacity in the concept of evidence accumulation. Specifically, these models suggest that the brain confronts the challenge of distilling reliable signals from the noise-contaminated sensory samples by integrating the samples over time,

averaging out the noise (Ratcliff et al., 2016). The instantiation of this simple idea in the Drift Diffusion Model provides an elegant account of behaviour on a wide variety of perceptual tasks (Ratcliff & Smith, 2004). However, the most powerful evidence of integration comes from recordings of sensorimotor neurons that exhibit firing patterns resembling the accumulation-to-bound dynamics described by the Drift Diffusion Model.

An inherent problem with the identification of a neural evidence accumulation signal is the need to dissociate that signal from other decision-related activity (Kelly & O'Connell, 2015). In the case of visual perceptual decisions, the excitation of photoreceptors sets off a cascade of neural events as the sensory data are funnelled through the cortical processing hierarchy. This activity is directly driven by the visual stimulus that is the subject of the perceptual decision, so much of the neural response is likely to be highly correlated with relevant stimulus features (e.g. motion direction, brightness, colour, etc.). Furthermore, some of that activity will represent the sensory evidence directly feeding the decision process. Therefore, an evidence accumulation signal, expressing the integral of the momentary evidence, needs to be disentangled from the backdrop of the highly-related activity elicited by that momentary evidence. Fortunately, several critical distinctions have been highlighted that can help in distinguishing sensory encoding signals from evidence accumulation signals.

Neural activity representing sensory encoding is expected to 1) scale with the momentary strength of the relevant stimulus feature; and 2) influence reaction time and choice independent of the physical stimulus (Kelly & O'Connell, 2015). Neural recording, microstimulation, and lesion studies have established that neurons in the middle temporal (MT) area fulfil these criteria when monkeys perform motion discrimination (Ditterich et al., 2003; Newsome & Paré, 1988; Platt, 2002). When the motion stimulus falls in the receptive field of an MT neuron, the sustained response is proportionate to the strength of motion coherence and dependent on its direction selectivity (Britten et al., 1992,1996; Newsome et al., 1989). Subsequent research has shown that the activity of these neurons predicts the monkey's choice, even in the absence of any physical evidence (Britten et al., 1996; Parker & Newsome, 1998). Furthermore, researchers successfully demonstrated that stimulating MT neurons with a response preference for a particular motion direction systematically biased the monkey's subsequent perceptual judgement in that direction (Salzman et al., 1990), proving that this neural activity feeds the decision process. Not only did the stimulation influence choice, but responses were faster when they were congruent with the preference of the stimulated neurons and slower when they conflicted with their direction tuning. Area MT encodes sensory evidence for motion stimuli, but the same pattern of results have been reported in other sensory areas for different perceptual tasks (Heekeren et al., 2008). For example, neurons in primary and secondary somatosensory cortex appear to feed decisions about vibrotactile stimuli (Hernández et al., 2000; Romo et al., 1998, 2000). This body of work demonstrates that sensory areas encode the instantaneous fluctuations of sensory samples in the evidence associated with the physical stimulus, but these populations do not deliberate on the collective representation of this evidence, instead they relay this ephemeral evidence to be assessed at subsequent processing stages (Mazurek et al., 2003).

In comparison to the activity underlying the encoding of sensory evidence, sequential sampling models predict that an evidence accumulation signal needs to demonstrate several additional characteristics: 1) the signal

should build over the course of the decision-making process and peak at the time of the response; 2) the rate of build-up should correlate with the strength of the sensory evidence; 3) perturbation of neural activity encoding the accumulated evidence should influence the decision process; and 4) inactivation of the region should significantly undermine task performance (Brody & Hanks, 2016; Shadlen & Kiani, 2013). The lateral intraparietal region (LIP) was identified as a natural candidate in the search for a region that might encode the decision variable in motion discrimination tasks, since it receives projections from area MT and is associated with the saccadic movements that are typically used to register responses in monkey studies.

A groundbreaking study from Shadlen and Newsome (1996) provided the first neurophysiological evidence of integration in LIP neurons. The authors observed that a subset of neurons whose receptive field was on the saccade target (displaced from the motion stimulus) exhibited a ramping in firing rates in advance of the saccadic response. While the build-up rate of this signal reflected the stimulus strength, the activity exhibited a stereotyped peak at response independent of motion coherence, resembling the crossing of a common decision bound (Roitman & Shadlen, 2002). In addition, this LIP activity predicted the monkey's choice on both correct and error response trials (Shadlen & Newsome, 2001). When the chosen response was in the opposite direction to the neuron's preference, the activity was found to decline at a rate proportional to the motion coherence, consistent with the Drift Diffusion Model's assertion that the decision variable represents the relative evidence for the competing alternatives (Gold & Shadlen, 2007). In tasks where the monkey has to wait for a

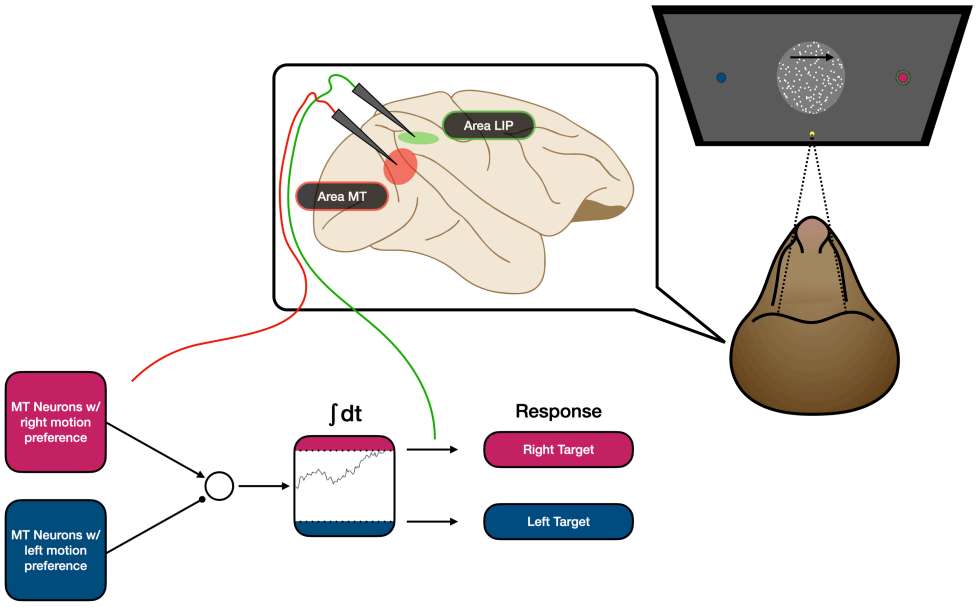


Figure 4.2. *Recording neural activity in areas MT and LIP while monkeys perform random dot motion discrimination (based on Shadlen & Newsome, 2001). The monkeys fixate on a central fixation point and indicate their decision with a saccade to the targets on the left or right of the motion stimulus. The green surround on the pink response target represents the receptive field of an LIP neuron. The activity recorded in MT maps to the sensory encoding stage of a simplified Drift Diffusion Model. LIP responses are thought to index a motor-related processing stage downstream of the original evidence accumulation stage (Ditterich, 2006; Katz et al., 2016).*

'go' signal before responding, these neurons sustain their activity during the delay, providing a temporal link between previously accumulated sensory evidence and delayed response execution (Kiani et al., 2008).

Gold and Shadlen (2000) surmised that since the monkey knows the response-mapping on the task following heavy training, the decision process could be achieved by translating accumulating motion evidence into oculomotor preparation for the associated response, in which case the dynamics of the decision process may be relayed to a downstream area more directly associated with this preparation of the saccade. In attempting to detect this activity, they interrupted the decision-making process with microstimulation of the frontal eye field (FEF) and simultaneous stimulus offset. This stimulation provoked a rapid eye movement before the voluntary saccade to the response target, and the trajectories of these initial movements left signatures of the evolving decision. The magnitude of the evoked eye movements was dependent on a combination of the motion coherence and the time that had elapsed before the stimulation. In addition, the direction of the movement predicted the monkey's upcoming choice, and the reliability of this prediction increased as stimulus coherence and viewing time increased, as would be expected of a decision variable diffusing towards a bound.

This evidence of integration is further bolstered by examination of the relationship between MT and LIP responses. Newsome et al. (1989) found that the motion sensitivity that could be achieved with just a pair of oppositely-tuned MT neurons responding to a 2 second motion stimulus rivalled the overall performance of the monkey itself. This was an astonishing result because it raises the question, what are all of the other MT neurons doing? In reality, activity in MT emerges from pools of neurons with diverse direction-selectivities, diminishing the 'neural sensitivity' in comparison to the perfectly oppositely-tuned pair (Shadlen et al., 1996) and subsequent modelling work suggests that the decision process may terminate long before the 2 seconds have elapsed (Kiani et al., 2008; Mazurek et al., 2003). Nonetheless, it is an important observation because it suggests that the decision process may read out estimated net motion calculated from the difference in spiking rates between these highly-sensitive, oppositely-tuned MT populations, consistent with the Drift Diffusion Model's contention that the decision variable encodes the relative evidence for the alternatives (Gold & Shadlen, 2001). While MT neurons preserve sustained responses to stable motion, LIP responses ramp over the course of stimulus presentation, consistent with the proposal that they express the integral of MT activity. However, when only random motion is presented, the LIP response is flat while the MT response is above zero. This would be expected if the LIP response represented the difference between the activity of pools of oppositely-tuned MT neurons whose receptive fields capture the motion stimulus (Shadlen & Newsome, 2001).

Mazurek et al. (2003) investigated if this proposed relationship between area MT and LIP could account for behavioural and neural data from a free response random dot motion task by simulating activity in these areas. The model's area MT was composed of two oppositely-tuned pools of motion-selective neurons, which responded in proportion to the motion coherence (based on neurophysiological recordings by Britten et al., 1993). Two pools of LIP neurons behaved as integrators, accumulating the relative sensory evidence in favour of either a right or left response towards a decision threshold. The model reproduced the pattern of decreasing reaction times with increasing stimulus coherence that is characteristic of behaviour on this task, suggesting

that this relationship emerges from faster rates of evidence accumulation at stronger coherences. In addition, the accumulation model accurately predicted the pattern of neural responses recorded in area LIP in response to a variety of coherence levels, with steeper slopes for stronger motion coherences in the evidence-locked signal and a stereotyped peak at response.

This body of evidence seems to position area LIP as the provisional neurophysiological locus of evidence accumulation. However, some have raised doubts about this proposal (Schall, 2019) and several studies have attempted to clarify this relationship by testing the causality of LIP activity in decision formation (Pisupati et al., 2016). Katz et al. (2016) argued that there is a danger of mistaking correlation with causation in assuming LIP encodes the decision variable. They sought to address this question by temporarily pharmacologically inactivating areas MT and LIP while monkeys performed a random dot motion task. The perceptual sensitivity of the monkeys was drastically reduced when MT was inactive, but surprisingly, interfering with area LIP had no significant impact on the monkeys' accuracy or their ability to integrate evidence. In contrast, Hanks et al. (2006) found that microstimulation of LIP neurons increased the proportion and speed of choices to the response target in their receptive field, slowing responses for the alternative choice. They concluded that LIP activity is part of the causal chain in decision formation, contributing to the representation of a decision variable for saccadic choices. Zhou and Freedman (2019) attempted to replicate Katz's findings, but they found inactivation of LIP neurons led to significant behavioural impairment when the neurons' receptive fields were aligned with either the motion stimulus or the saccadic response target. Modelling indicated that the former stymied the rate of evidence integration, while the latter biased the decision process away from the choice associated with the inactivated neurons. Consistent with this finding, Yates et al. (2017) reported that LIP responses could not be explained solely by the integration of MT responses to pulses of motion. Instead, LIP activity was characterised as a mixture of sensory integration and response preparation buildup signals. Others have even challenged the characterisation of LIP activity as 'ramping', arguing that the supposed ramping is an artifact of trial averaging and that the activity is better described as comprising discrete steps (Latimer et al., 2015; Zoltowski et al., 2019); however, this argument has been strongly refuted by others (Shadlen et al., 2016).

Together, these results indicate that it is unlikely that LIP passively receives decision-related information relayed from an upstream area. Instead, Zhou and Freedman (2019) argue that LIP has a central role in translating sensory evidence into saccadic decisions. It is also possible that LIP is part of a constellation of decision-related modules, recruited flexibly depending on the response modality, which offer sufficient redundancy to compensate for its inactivation. Indeed, converging lines of neurophysiological research in monkeys have provided strong evidence that the decision state is readily detectable in a distributed network of regions associated with the preparation of a motor response, including the superior colliculus (Ratcliff et al., 2007), and FEF (Ding & Gold, 2012; Gold & Shadlen, 2000) for oculomotor responses, dorsal premotor cortex and the medial intraparietal area for tasks that demand reaching responses (Cisek & Kalaska, 2005; de Lafuente et al., 2015), and somatosensory and ventral premotor cortex for vibrotactile decisions (Romo et al., 2000, 2004). As Pisupati et al. (2016) point out, just because LIP may not be the originator of these decision signals, does not mean that activity recorded from LIP cannot be informative about computational dynamics of the causal source

or sources. Recordings of LIP neurons still represent powerful evidence of integration, and have offered valuable insight into the flexibility of the neural circuits underlying decision-making in different contexts (O'Connell & Kelly, 2021).

Neurophysiological Evidence of Accumulation in Humans

These exciting advances made in invasive neurophysiological research spurred efforts to profile the same decision-making process in humans, albeit with non-invasive methods lacking the spatial and temporal resolution of direct neural recordings (O'Connell et al., 2012). In the last decade, a confluence of methodological advances in signal processing, computational modelling, and task design made this possible, allowing human research to complement and expand on the foundations laid in monkey neurophysiology (O'Connell & Kelly, 2021).

In 2012, O'Connell et al. reported the detection of an ERP component which reflected the dynamics of the evidence accumulation process. The authors asked participants to detect gradual changes in a contrast stimulus over long viewing durations. The centroparietal positivity (CPP) exhibited a slow build-up over the course of the decision and reached a stereotyped peak predicting the response. Perturbing the contrast target sequence produced lasting deflections in the CPP trajectory, consistent with an evidence integration signal. Critically, the CPP's gradual build-up scales with the strength of the sensory evidence (Steinemann et al., 2018), in a manner consistent with the accumulation of sensory samples (Kelly & O'Connell, 2013). As discussed above, this pattern of activity has been observed in a distributed network of oculomotor regions and subsequent research on manual response tasks identified the same evidence-dependent, choice-predictive dynamics in human electrophysiological motor preparation signals, like the desynchronisation of beta band activity over premotor cortex (Donner et al., 2009; Kubanek et al., 2013; O'Connell et al., 2012). However, these premotor decision signals are inherently tethered to the response modality; regions like LIP and FEF only represent the decision state for saccadic tasks, beta desynchronisation is only observed for tasks requiring manual responses (O'Connell et al., 2012), and if the stimulus-response mapping is withheld until after evidence presentation, beta-activity does not recapitulate decision formation (Twomey et al., 2015). The CPP is distinguished from all of these effector-selective signatures of the decision process by its remarkable versatility.

The CPP exhibits the same evidence-dependent build-up dynamics independent of the sensory modality or the relevant stimulus feature (O'Connell & Kelly, 2021), and this remains true even when no response is required or when the stimulus-response mapping is not known (O'Connell et al., 2012; Twomey et al., 2015). Unlike motor preparation signals, the positive polarity and topography of the CPP is the same for all perceptual decisions. The CPP also diverges from motor preparation signals in its response to speed pressure. Beta-band activity adjusts to speed pressure with a shift towards greater baseline motor preparation, while the CPP exhibits a constant starting point. Both signals reach a peak at response, but beta activity converges on a stereotyped threshold independent of decision urgency or reaction time, while the CPP exhibits a reduced amplitude at

response for decisions with narrow response deadlines (Kelly et al., 2021; Steinemann et al., 2018). Computational modelling of these dynamics indicates that motor preparation signals encode a representation of both the accumulated evidence and strategic adjustments like decision urgency (accounting for the elevated baseline), but the CPP solely represents the cumulative evidence component of the decision process (Kelly et al., 2021; Steinemann et al., 2018). These observations are in close accord with the dynamics predicted by a sequential sampling model with a motor-level urgency component that narrows the decision bounds as time elapses, reducing the quantity of evidence needed to terminate the decision.

Thus, the CPP appears to encode a pure, supramodal, motor-independent representation of cumulative evidence (O'Connell & Kelly, 2021), providing opportunities to non-invasively profile the dynamics of decision model parameters in tasks ranging from simple vibrotactile discriminations (Herding et al., 2019) to complex perceptual judgements like facial recognition (van Vugt et al., 2019). While the CPP offers a unique perspective on evidence accumulation in humans, its neural origins are largely unknown and there is some ambiguity about what exactly the CPP represents (O'Connell & Kelly, 2021). Kelly and O'Connell (2015) suggest that the CPP may emerge from intermingled pools of parietal neurons behaving as accumulators, similar to LIP neurons, and racing against each other as evidence accumulates. In such an arrangement, it is unclear if the CPP represents the total cumulative evidence for all decision alternatives or if mutual inhibition between the accumulator pools may result in an expression of the same quantity as the decision variable in the Drift Diffusion Model (i.e. the relative evidence favouring the current perceptual hypothesis). Distinguishing between these accounts was one of the goals of Chapter 5.

Dissociating Evidence Accumulation and Response Preparation

Traditionally, cognitive theories of decision-making assume that the decision process precedes action planning in a serial arrangement. However, Cisek (2007) points out that a serial decision-making system does not make sense for animals that evolved in a capricious and hostile environment where opportunity and danger demand rapid responses at short notice. According to the Affordance Competition Hypothesis (Cisek, 2007), the accumulation of sensory evidence is used to specify and adjudicate between multiple competing potential actions. This produces a more fluid, embodied decision-making, where the objective is not purely to arrive at an accurate depiction of one's surroundings, but to mediate adaptive interaction with the environment (Cisek, 2007). Regardless of the veracity of this particular account, attempts to bridge neurophysiology and behaviour have provided grounds for rejecting a serial architecture of sensory, decision-making, and action planning systems based on the observation that the cortex is not neatly arranged to segregate sensory, decision, and motor activity (Cisek et al., 2009; Goodale & Milner, 1992; Heekeren et al., 2008). Instead, neurophysiological recordings of decision signals are consistent with a continuous updating of motor planning throughout deliberation, such that the state of the decision variable can be read out in response-related areas like FEF

(Ding & Gold, 2012; Gold & Shadlen, 2000) and LIP (Shadlen & Newsome, 2001) while the decision is still evolving.

This issue raises important questions about the CPP. If the CPP represents an abstract decision variable upstream of motor preparation, it should encode the decision state before continuously relaying that information to the downstream areas associated with response preparation. This means that, as new sensory evidence is fed into the decision process, it should be detectable in the trajectory of the CPP in advance of any effector-specific signals. One such signal is the lateralised readiness potential (LRP), which is an ERP component that emerges in the difference in contralateral and ipsilateral activity over motor cortex in the lead up to a manual response (Ikeda & Shibasaki, 1992; Rinkenauer et al., 2004; van Vugt et al., 2014). Like beta activity, It has been demonstrated that the LRP reflects the accumulated evidence as predicted by sequential sampling models; building up over the course of the decisions requiring a manual response, with a slope proportionate to the strength of the evidence, and reaching a stereotyped peak at response (Kelly & O'Connell, 2013; Polanía et al., 2014). However, beta-band activity suffers from an impoverished temporal resolution resulting from the time-frequency techniques used to extract the signal from EEG. In contrast, the LRP is the perfect candidate to assess the relative timing of CPP and response preparation signals because it shares the same millisecond temporal resolution as the CPP. Kelly and O'Connell (2013) reported that the earliest period in which the buildup of the CPP was reliably influenced by motion coherence occurred approximately 150 ms before the LRP. The result points to the possibility that cumulative sensory evidence may be encoded in a supramodal decision variable represented by the CPP before being conveyed to motor areas preparing the specific decision-reporting actions.

To our knowledge, this is the only example of such an analysis. The experiment described in Chapter 5 was designed to replicate this analysis using a more sensitive measurement of the relative latencies of these two signals. While the analysis reported by Kelly and O'Connell (2013) relied on the response to the onset of evidence, this study incorporates precisely timed bursts of sensory evidence with unpredictable onsets facilitating an analysis of the timing of CPP and LRP responses to the pulses. An advantage of this approach is that the pulses could perturb evidence in opposite directions, allowing for timing analyses based on strong modulations of signal direction as well as build-up rate.

The Value of Computational Modelling in Perceptual Research

The Drift Diffusion Model's success is attributable to its capacity to accommodate both behavioural data from myriad perceptual tasks and groundbreaking neurophysiological observations of decision process correlates under a unifying theoretical framework (Gold & Shadlen, 2007; Kelly & O'Connell, 2015; Ratcliff et al., 2016). Indeed, formal decision models are critical for bridging the gap between behaviour and neurophysiology, serving as an intermediate layer of analysis, mapping candidate computational mechanisms to their observable output (Mulder et al., 2014; O'Connell et al., 2018). Interrogating the dynamics of this theoretical nexus in

different conditions has proven a fruitful strategy, as researchers have begun to decompose behavioural adjustments to changing task demands into latent psychological processes that were entirely inaccessible without mathematical modelling (Forstmann et al., 2016). As a more sophisticated understanding of the mechanisms of perceptual decision-making emerged from this work, efforts to profile the elusive relationships between model parameters and neural activity have forged new and exciting avenues for progress, including the identification of novel neural markers of development, ageing, and disease (Forstmann et al., 2016; Ratcliff et al., 2016; Ratcliff & Van Dongen, 2011; van Ravenzwaaij et al., 2012).

One example of such an adjustment that has received a lot of attention is the speed-accuracy tradeoff, where participants instructed to make rapid decisions exhibit lower accuracy and faster errors than those provided with generous response deadlines. Converging evidence suggests that the effect emerges, at least in part, from the lowering of decision bounds as time pressure increases (Forstmann et al., 2008; O'Connell et al., 2018; Steinemann et al., 2018), with lower bounds reflecting a less cautious decision strategy and higher bounds reducing the chances of an error, but also increasing the time taken to make the decision. Several important avenues of investigation spring from this model-based approach: correlational analyses can be applied to find where the bound adjustment is computed in the brain (Forstmann et al., 2008), research on age-related cognitive decline may develop screening measures and interventions based on individual differences in speed-accuracy compensations in different populations (Dully et al., 2018), more subtle complementary adjustments may be revealed in paradigms tailored to capture the bound adjustment (Steinemann et al., 2018), and signatures of the neurophysiological implementation of the adjustment can be identified based on the predictions of the bound adjustment account (Hanks et al., 2014).

This example illustrates that the value of computational models like the Drift Diffusion Model is not necessarily in faithful replication of the underlying neural processing, but primarily in the construction of a meaningful framework for generating and addressing fundamental questions that could not be answered by either behavioural or neurophysiological data alone (Mulder et al., 2014). However, just as mathematical models can aid the interpretation of empirical phenomena, neurophysiological data can also inform the construction of models and test key model predictions. For instance, there is a long-standing controversy over the role of dynamic urgency in the speed-accuracy trade-off (Boehm et al., 2020; Ratcliff et al., 2016). Most behavioural modelling studies have endorsed a static bound adjustment (Hawkins et al., 2015), but neurophysiological investigations have consistently reported that neural decision signals in effector-selective areas exhibit evidence-independent build-up components consistent with dynamic urgency, suggesting that collapsing bounds help to accelerate responses under time pressure (Churchland et al., 2008; Hanks et al., 2014; Murphy et al., 2016; Steinemann et al., 2018; Thura & Cisek, 2016). By constraining several parameters with neurophysiological correlates, Kelly et al. (2021) found that a model with urgency could better account for their data than the Drift Diffusion Model. Furthermore, the neurally-informed model revealed several novel speed-accuracy adjustments that were not only absent from, but directly contradicted by the Drift Diffusion Model fit.

References

Boehm, U., van Maanen, L., Evans, N. J., Brown, S. D., & Wagenmakers, E.-J. (2020). A theoretical analysis of the reward rate optimality of collapsing decision criteria. *Attention, Perception & Psychophysics*, *82*, 1520–1534.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience*, *13*, 87–100.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *The Journal of Neuroscience*, *12*, 4745–4765.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience*, *10*, 1157–1169.

Brody, C. D., & Hanks, T. D. (2016). Neural underpinnings of the evidence accumulator. *Current Opinion in Neurobiology*, *37*, 149–157.

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*, 693–702.

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron*, *45*, 801–814.

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: the urgency-gating model. *The Journal of Neuroscience*, *29*, 11560–11571.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*, 1585–1599.

de Lafuente, V., Jazayeri, M., & Shadlen, M. N. (2015). Representation of accumulating evidence for a decision in two parietal areas. *The Journal of Neuroscience*, *35*, 4306–4318.

Ding, L., & Gold, J. I. (2012). Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cerebral Cortex*, *22*, 1052–1067.

Ditterich, J., Mazurek, M. E., & Shadlen, M. N. (2003). Microstimulation of visual cortex affects the speed of perceptual decisions. *Nature Neuroscience*, *6*, 891–898.

Donner, T. H., Siegel, M., Fries, P., & Engel, A. K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Current Biology*, *19*, 1581–1585.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *The Journal of Neuroscience*, *32*, 3612–3628.

Dully, J., McGovern, D. P., & O'Connell, R. G. (2018). The impact of natural aging on computational and neural indices of perceptual decision making: A review. *Behavioural Brain Research*, *355*, 48–55.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 17538–17542.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. *Annual Review of Psychology*, *67*, 641–666.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

Gold, J. I., & Stocker, A. A. (2017). Visual Decision-Making in an Uncertain and Dynamic World. *Annual Review of Vision Science*, *3*, 227–250.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*, 390–394.

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, *5*, 10–16.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20–25.

Hanks, T. D., Kiani, R., & Shadlen, M. N. (2014). A neural mechanism of speed-accuracy tradeoff in macaque area LIP. ELife, 3.

Hanks, T. D., Ditterich, J., & Shadlen, M. N. (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nature Neuroscience*, *9*, 682–689.

Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *The Journal of Neuroscience*, *31*, 6339–6352.

Hanks, T. D., & Summerfield, C. (2017). Perceptual decision making in rodents, monkeys, and humans. *Neuron*, *93*, 15–31.

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *The Journal of Neuroscience*, *35*, 2476–2484.

Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews. Neuroscience*, *9*, 467–479.

Herding, J., Ludwig, S., von Lautz, A., Spitzer, B., & Blankenburg, F. (2019). Centro-parietal EEG potentials index subjective evidence and confidence during perceptual decision making. *Neuroimage*, *201*, 116011.

Hernández, A., Zainos, A., & Romo, R. (2000). Neuronal correlates of sensory discrimination in the somatosensory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 6191–6196.

Ikeda, A., & Shibasaki, H. (1992). Invasive recording of movement-related cortical potentials in humans. *Journal of Clinical Neurophysiology*, *9*, 509–520.

Katz, L. N., Yates, J. L., Pillow, J. W., & Huk, A. C. (2016). Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature*, *535*, 285–288.

Kelly, S. P., Corbett, E. A., & O'Connell, R. G. (2021). Neurocomputational mechanisms of prior-informed perceptual decision-making in humans. *Nature Human Behaviour*, *5*, 467–481.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *The Journal of Neuroscience*, *33*, 19434–19441.

Kelly, S. P., & O'Connell, R. G. (2015). The neural processes underlying perceptual decision making in humans: recent progress and future directions. *Journal of Physiology, Paris*, *109*, 27–37.

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *The Journal of Neuroscience*, *28*, 3017–3029.

Kubanek, J., Snyder, L. H., Brunton, B. W., Brody, C. D., & Schalk, G. (2013). A low-frequency oscillatory neural signal in humans encodes a developing decision variable. *Neuroimage*, *83*, 795–808.

Latimer, K. W., Yates, J. L., Meister, M. L. R., Huk, A. C., & Pillow, J. W. (2015). NEURONAL MODELING. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*, *349*, 184–187.

Martinez-Rodriguez, L. A., Corbett, E. A., O'Connell, R. G., & Kelly, S. P. (2020). Neurally-Informed Modelling of Static and Dynamic Decision Biases. *Research Gate*.

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex*, *13*, 1257–1269.

Mulder, M. J., van Maanen, L., & Forstmann, B. U. (2014). Perceptual decision neurosciences - a model-based review. *Neuroscience*, *277*, 872–884.

Murphy, P. R., Boonstra, E., & Nieuwenhuis, S. (2016). Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nature Communications*, *7*, 13526.

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, *341*, 52–54.

Newsome, W. T., & Paré, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *The Journal of Neuroscience*, *8*, 2201–2211.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*, 1729–1735.

O'Connell, R. G., & Kelly, S. P. (2021). Neurophysiology of Human Perceptual Decision-Making. *Annual Review of Neuroscience*, *44*, 495–516.

O'Connell, R. G., Shadlen, M. N., Wong-Lin, K., & Kelly, S. P. (2018). Bridging Neural and Computational Viewpoints on Perceptual Decision-Making. *Trends in Neurosciences*, *41*, 838–852.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. *Annual Review of Neuroscience*, *21*, 227–277.

Pisupati, S., Chartarifsky, L., & Churchland, A. K. (2016). Decision Activity in Parietal Cortex - Leader or Follower? *Trends in Cognitive Sciences*, *20*, 788–789.

Platt, M. L. (2002). Neural correlates of decisions. *Current Opinion in Neurobiology*, *12*, 141–148.

Polanía, R., Krajbich, I., Grueschow, M., & Ruff, C. C. (2014). Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. *Neuron*, *82*, 709–720.

Ratcliff, R., Hasegawa, Y. T., Hasegawa, R. P., Smith, P. L., & Segraves, M. A. (2007). Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *Journal of Neurophysiology*, *97*, 1756–1774.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: current issues and history. *Trends in Cognitive Sciences*, *20*, 260–281.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367.

Ratcliff, R., & Van Dongen, H. P. A. (2011). Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 11285–11290.

Rinkenauer, G., Osman, A., Ulrich, R., Muller-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General*, *133*, 261–282.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of Neuroscience*, *22*, 9475–9489.

Romo, R., Hernández, A., Zainos, A., Brody, C. D., & Lemus, L. (2000). Sensing without touching: psychophysical performance based on cortical microstimulation. *Neuron*, *26*, 273–278.

Romo, R., Hernández, A., Zainos, A., & Salinas, E. (1998). Somatosensory discrimination based on cortical microstimulation. *Nature*, *392*, 387–390.

Romo, R., Hernández, A., & Zainos, A. (2004). Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron*, *41*, 165–173.

Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, *346*, 174–177.

Schall, J. D. (2019). Accumulators, neurons, and response time. *Trends in Neurosciences*, *42*, 848–860.

Shadlen, M. N., Kiani, R., Newsome, W. T., Gold, J. I., Wolpert, D. M., Zylberberg, A., … Roitman, J. (2016). Comment on "Single-trial spike trains in parietal cortex reveal discrete steps during decision-making". *Science*, *351*, 1406.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *The Journal of Neuroscience*, *16*, 1486–1510.

Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, *80*, 791–806.

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 628–633.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*, 1916–1936.

Steinemann, N. A., O'Connell, R. G., & Kelly, S. P. (2018). Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nature Communications*, *9*, 3627.

Summerfield, C., & Blangero, A. (2017). Perceptual Decision-Making. In *Decision Neuroscience* (pp. 149–162). Elsevier.

Thura, D., & Cisek, P. (2016). Modulation of Premotor and Primary Motor Cortical Activity during Volitional Adjustments of Speed-Accuracy Trade-Offs. *The Journal of Neuroscience*, *36*, 938–956.

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *The European Journal of Neuroscience*, *42*, 1636–1643.

van Ravenzwaaij, D., Dutilh, G., & Wagenmakers, E.-J. (2012). A diffusion model decomposition of the effects of alcohol on perceptual decision making. *Psychopharmacology*, *219*, 1017–1025.

van Vugt, M. K., Beulen, M. A., & Taatgen, N. A. (2019). Relation between centro-parietal positivity and diffusion model parameters in both perceptual and memory-based decision making. *Brain Research*, *1715*, 1–12.

van Vugt, M. K., Simen, P., Nystrom, L., Holmes, P., & Cohen, J. D. (2014). Lateralized readiness potentials reveal properties of a neural mechanism for implementing a decision threshold. *Plos One*, *9*, e90943.

Yates, J. L., Park, I. M., Katz, L. N., Pillow, J. W., & Huk, A. C. (2017). Functional dissection of signal and noise in MT and LIP during decision-making. *Nature Neuroscience*, *20*, 1285–1292.

Zhou, Y., & Freedman, D. J. (2019). Posterior parietal cortex plays a causal role in perceptual and categorical decisions. *Science*, *365*, 180–185.

Zoltowski, D. M., Latimer, K. W., Yates, J. L., Huk, A. C., & Pillow, J. W. (2019). Discrete Stepping and Nonlinear Ramping Dynamics Underlie Spiking Responses of LIP Neurons during Decision-Making. *Neuron*, *102*, 1249-1258.e10.

5.

The Role of Evidence Accumulation in Perceptual Decisions About Low-Noise Stimuli

The central problem addressed in this thesis is how a reliable, useful perceptual experience can be constructed from an underspecified sensory stream polluted with internal and external sources of noise. Sequential sampling models propose that one solution may be to simply collect more data. If the objective is to isolate the signal from its contaminants, the law of large numbers dictates that the estimation will grow increasingly accurate in proportion to the square root of the samples. In other words, initially ambiguous sensory information can be rendered interpretable when sampled over longer periods of time. As discussed in Chapter 4, mathematical models that hinge on the proposal that perceptual decisions are accomplished through evidence accumulation have been critical to the progress that has been made in this field (Summerfield & Blangero, 2017). In parallel with their development, extensive investigations of behavioural and neurophysiological data have sought to determine the degree to which participants rely on evidence accumulation when confronted with a noisy stimulus like random dot motion (Gold & Shadlen, 2007). Some of this work has raised questions about whether the default assumption of integration is justified. A key area of

debate now pertains to the possibility that novel model variants eschewing extended evidence integration could provide a more compelling account of perceptual decision-making.

Evidence accumulation models make a number of predictions about how behaviour should vary across different experimental conditions. If evidence integration is being used to minimise noise, the decision process should be more efficient with longer stimulus exposure and less noisy stimuli. Indeed, in free-response paradigms, when subjects are presented with strong coherent motion, reaction times are fast and accuracy is high; when they are presented with weak motion, reaction times slow down and accuracy falls (Ratcliff & McKoon, 2008; Roitman & Shadlen, 2002). Furthermore, when the stimulus duration is controlled by the experimenter, choice accuracy increases as the stimulus is presented for increasing durations (Britten et al., 1992; de Lafuente et al., 2015; Gold & Shadlen, 2000; Kiani et al., 2008). The most prominent example of a sequential sampling model, the Drift Diffusion Model (Ratcliff et al., 2016), assumes perfect integration on typical decision-making tasks (lasting ~2 seconds); that is, every sample of sensory evidence continues to exert an influence on the trajectory of the decision variable for the duration of deliberation. Maximising the number of samples integrated by the decision variable minimises the average noise in the signal estimate, optimising perceptual sensitivity when the decision bound is crossed (Wald & Wolfowitz, 1948). Indeed, in tasks where the stimulus duration varies from trial to trial, the decision appears to be shaped by samples of evidence across the entire stimulus presentation (Brunton et al., 2013), and there is evidence that perceptual sensitivity in rodents, monkeys, and humans improves with longer stimulus duration at a rate predicted by perfect integration (Brunton et al., 2013; Huk & Shadlen, 2005; Kiani et al., 2008). The evidence presented in Chapter 4 heavily favours an integration-based perceptual decision-making account: evidence accumulation accounts for behaviour on a wide variety of tasks (Ratcliff et al., 2016), neurons exhibiting the same accumulation-to-bound dynamics predicted by the Drift Diffusion Model have been identified in multiple sensorimotor areas (Gold & Shadlen, 2007), and an abstract, effector-independent electrophysiological signature of the same process has been found in humans (O'Connell & Kelly, 2021). However, this research has focussed on a narrow range of perceptual tasks, in particular random dot motion discrimination, and recent computational modelling work has highlighted that alternative model variants with little or no integration can often provide equally good fits to perceptual choice data (Carland et al., 2016; Cisek et al., 2009; Thura et al., 2012).

Challenges in Verifying Integration Accounts in Perceptual Decision-Making

There is substantial ongoing debate about the timescale over which integration operates and some models propose that there is no evidence integration at all. Like the Drift Diffusion Model, 'Extrema Detection' (Watson, 1979) suggests that evidence is sequentially sampled over time, but instead of perfectly integrating evidence towards a decision boundary (Figure 5.1.A), the Extrema Detection decision process immediately discards all evidence that does not exceed the decision threshold (Figure 5.1.B). When an 'extremum' is detected the decision process is terminated. Perfect integration strategies can also be compared to leaky models that

'forget' sensory samples after some period of time; models with strong leak represent only the most recent evidence, while models with little leak only discard the earliest samples. For instance, the Leaky Competing Accumulator Model (Usher & McClelland, 2001) proposes that the decision process is mediated by competition between mutually-inhibitory leaky accumulators that race to threshold. Alternatively, in the Urgency Gating Model (Cisek et al., 2009), the decision variable is a low-pass filtered estimate of momentary evidence multiplied by an urgency signal, which serves to promote a response before the response deadline. In each case, the time constant (i.e. the period with no leak) is typically 100-200 ms (Carland et al., 2016; Cisek et al., 2009; Thura et al., 2012; Usher & McClelland, 2001), so although they are not equivalent to the instantaneous comparisons of Extrema Detection, these strategies are far from the perfect integration proposed by the Drift Diffusion Model. In this sense, integration can be viewed as a continuum, ranging from accounts that posit perfect or near-perfect integration (e.g. Drift Diffusion Model) to those that propose narrow windows of integration (e.g. the Urgency Gating Model) or no integration at all (e.g. Extrema Detection). Throughout the chapter the term 'extreme leak' is used to refer to decision-making strategies that assume either no integration (e.g. Extrema Detection) or integration with very restricted time constants (e.g. the Urgency Gating Model).

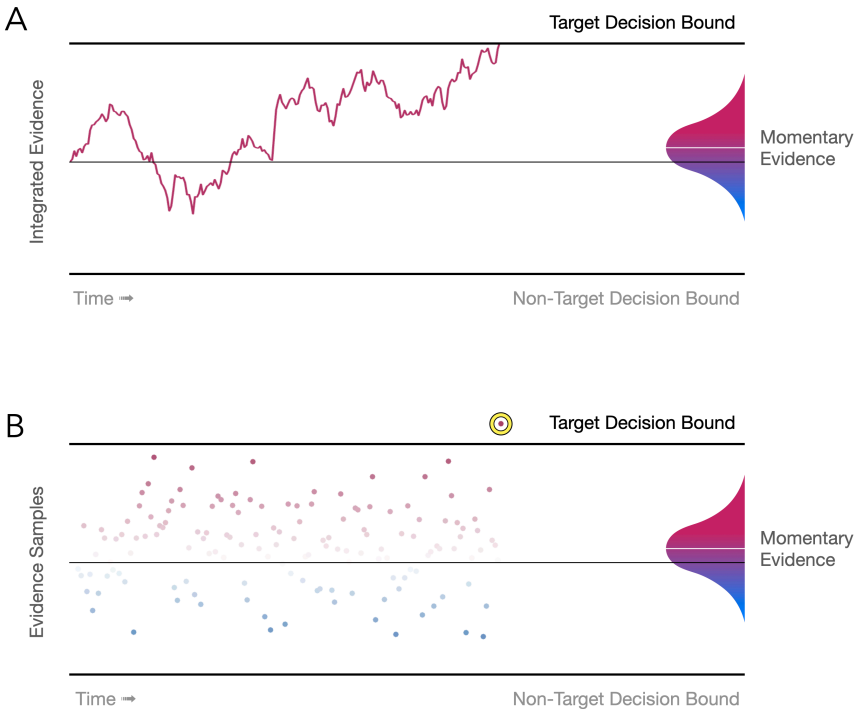


Figure 5.1. *Drift Diffusion and Extrema Detection accounts of two-alternative perceptual decision-making. In each model, noisy evidence is sequentially sampled from a distribution of values representing the encoded sensory information. If the stimulus is above the psychophysical threshold, the distribution is biased towards the relevant parameter of the physical stimulus (e.g. rightwards motion or left-tilted orientation; note the white line representing the median value). A) According to the Drift Diffusion model, this noisy evidence is accumulated with no leak until a decision bound is reached and a response is initiated. B) The Extrema Detection model does not integrate evidence. Instead each sample is independently evaluated and, if it fails to reach the decision bound, it is immediately discarded. The decision process stops when an extremum (a sample that crosses this threshold) is detected. In Stine et al.'s model, if an extremum is not detected before the end of the trial, the decision is a coin toss.*

Ditterich (2006) sought to clarify the distinctions between integration and extreme leak, comparing the capacity of variants of the Drift Diffusion Model and Extrema Detection to account for behavioural and neurophysiological data collected from macaques performing a free response random dot motion discrimination task (Roitman & Shadlen, 2002). He found that both Extrema Detection and the Drift Diffusion Model were able to account for the behavioural data, producing identical psychometric functions. However, three telling signatures of the inadequacy of Extrema Detection emerged on closer inspection. First, the extreme leak models afforded very little time to the decision process, estimating non-decision time to be twice as long as the integration models. Second, the extreme leak models were amongst the poorest fits to the neural recordings from area LIP, doubling the estimated deviation from LIP trajectories of the best-fitting integration model. Finally, the signal-to-noise ratio required for the extreme leak model fits was estimated to be ~10 times better than the signal-to-noise ratio of sensory encoding amongst MT neurons, based on estimates derived from actual firing rates (Britten et al., 1993; Zohary et al., 1994). On the other hand, the integration model estimates were perfectly in line with the physiological estimate, and were deemed biologically plausible.

Stine et al. (2020) compared Extrema Detection and the Drift Diffusion Model by simulating behavioural data in fixed duration, variable duration, and free response random dot motion paradigms. For fixed stimulus durations, Extrema Detection mimicked the behavioural features of the Drift Diffusion Model, producing a psychophysical function with the same slope and indistinguishable temporal weighting profiles. An integration strategy was expected to distinguish itself in varying duration trials, since sensitivity to the underlying signal increases as the influence of noise is reduced (Brunton et al., 2013; Kiani et al., 2008). However, Stine et al. found that Extrema Detection also showed improved accuracy with increased stimulus durations. This was a consequence of a random guessing rule for trials where an extremum has yet to be detected when the evidence offsets. Since any strategy based on the evidence offers better performance than chance, outcomes are better for trials that stave off this relapse to 50% accuracy for longer. In addition, the Drift Diffusion Model could nicely fit the data generated by the Extrema Detection model in this task, underscoring the theme that non-integration strategies can easily masquerade as integration. Finally, comparing the simulations for the free response paradigm, again, Extrema Detection replicated behavioural trends normally considered signatures of integration on these tasks and each model was perfectly capable of fitting the reaction time and choice accuracy data generated by the other model.

However, Stine et al. identified that in achieving all of these fits, the Drift Diffusion Model and the Extrema Detection consistently differed on the same parameters highlighted by Ditterich (2006): signal-to-noise ratio and non-decision time. They suggested that the mimicry of integration in the fixed stimulus duration and free response tasks was attributable to flexibility in the signal-to-noise ratio and non-decision time parameters, respectively. To test this, they conducted a random dot motion experiment with human participants that interleaved variable duration and free response trials and constrained the parameters of interest using an empirically derived estimate of non-decision time and a signal-to-noise ratio estimated on variable duration trials. With the constrained parameters, Extrema Detection was not capable of accounting for the data on free

response trials, but the integration model successfully reproduced participants' psychometric functions and reaction time distributions.

Although these studies still support the role of evidence integration on tasks of this kind, they clearly demonstrate the substantial overlap in the behavioural patterns predicted by these distinct classes of model and the significant risk of misattributing evidence accumulation to patterns of behavioural data. However, Stine et al. (2020) do point out that there are other means of testing for integration aside from detailed behavioural modelling (e.g. Glickman & Usher, 2019; Waskom & Kiani, 2018). For example, Huk and Shadlen (2005) reported that perturbing the sensory evidence, by incorporating brief pulses in the motion stimulus, reliably influenced subsequent response times and produced persistent deflections in the trajectory of LIP activity that predicted those changes. Although the pulses only lasted 100 ms, their influence on reaction times and choices lasted for up to 900 ms. Similarly, the pulse effect on LIP activity was observed from 225 ms up to 800 ms after pulse onset, lingering for much longer than would be expected of an extreme leak process.

On the other hand, Carland et al. (2016) reported that human performance on a random dot motion task, incorporating brief pulses of positive evidence, was better accounted for by the Urgency Gating Model than the Drift Diffusion Model. They observed that early pulse effects were weakened as reaction times increased in blocks where evidence sometimes fluctuated within the trial, but early pulses were more influential on blocks where the evidence was always stable. This is consistent with the idea that when evidence was stable, late pulses were more often ignored because the decision variable was proximate to, or had already crossed, the decision bound (Mazurek et al., 2003); when participants delayed commitment to account for the variable evidence, the later pulses became more influential and the early pulse of evidence leaked out of the decision process. This pattern is not easily reconciled with the perfect integration assumed in the Drift Diffusion Model. However, using a motion energy analysis, Kiani et al. (2008) showed that random fluctuations in the sensory evidence were less influential on monkeys' choices at later stages of random dot motion trials with variable durations (>700 ms) that exceeded the time constant of leak models. Since these fluctuations are evenly spread across evidence presentation, proximity to the decision bound cannot explain this trend. To follow-up, Kiani et al. incorporated pulses into their variable duration random dot motion task and found that the choice bias exerted by the pulse diminished as onset time increased, with later pulses often ignored because the decision variable was proximate to, or had already crossed, the decision bound. There are a number of important differences between the studies that could explain why the findings of these studies appear to disagree on the role of extended evidence accumulation. First, Carland et al. allowed participants to respond at any time, but Kiani et al. made subjects wait for a 'go' signal. Carland et al. also compared behaviour in two conditions, one where all trials had the same difficulty and another where the difficulty sometimes fluctuated during the trial. This difference in task structure may have led to discrepancies in decision strategy. Additionally, while Carland et al. studied human behaviour, Kiani et al. recorded LIP activity in awake behaving monkeys, so the level of training on the task differed substantially between these studies. Finally, Carland et al. subdivided participants into fast and slow groups and the pattern of behaviour predicted by the Urgency Gating Model was only observed in the fast group.

Low-Noise Perceptual Tasks

Research probing the role of integration in perceptual decision-making has centred on a small subset of noisy discrimination tasks and it is unclear to what extent the principles and processes uncovered in this work extend to other scenarios. Several studies have provided evidence to suggest that participants can adapt their window of integration to the statistics of a noisy stimulus (Ghose, 2006; Ossmy et al., 2013) and in simple detection tasks, it is generally assumed that the visual system uses alternative decision-making strategies, like Extrema Detection (Robson & Graham, 1981; Sachs et al., 1971; Watson, 1979). Therefore, characterising perceptual decision-making in a greater variety of paradigms is an important step toward establishing how broadly evidence accumulation is applied.

One common scenario that is rarely considered in the field is perceptual choice based on low-noise stimuli. Most of the research discussed in this chapter, and indeed in perceptual decision-making research generally, has used the random dot motion task. This is the prototypical noisy evidence stimulus, since the signal (net motion energy) is embedded amongst noise (random motion). According to Carland et al. (2016), when an evidence signal is stable over time, as in typical versions of the random dot motion task, the reduction in noise achieved by integrating evidence will diminish with each successive sample. Once the noise has been filtered out by accumulating a sufficient number of samples, there is no point in continuing to integrate, especially when the representation of accumulated evidence comes at a metabolic cost (Drugowitsch et al., 2012). In the case of low-noise stimuli, this point should be reached far more rapidly as the primary source of noise to be minimised is associated with sensory encoding itself (Manwani & Koch, 2001; Parker & Newsome, 1998). Thus, it is far from certain a priori that evidence accumulation processes would benefit perceptual performance on low-noise tasks.

Relatively few studies have examined the role of integration in low-noise tasks. In one example, Ratcliff and Smith (2010) found that the Drift Diffusion Model perfectly accounted for behavioural performance on a brightness discrimination task with constant evidence and no stimulus noise, but they did not compare these fits to those of an extreme leak model. In another example, Brunton et al. (2013) modelled data from rat and human subjects tasked with determining which of two auditory streams contained a higher number of clicks or which orientation was dominant in a stream of oriented bars. The estimated time-constant of integration was as long as the stimulus duration, indicating that subjects perfectly integrated the evidence. Waskom and Kiani (2018) presented participants with multiple brief exposures to a contrast stimulus in trials lasting 10 seconds on average, and participants were asked to determine whether the samples came from a high- or low-contrast distribution. While an Extrema Detection account mimicked features of integration, overall behavioural performance was only consistent with an integration strategy with no leak. However, it is not known whether decision-making on expanded judgements like this is comparable to the more rapid perceptual decisions commonly studied using paradigms like random dot motion discrimination.

O'Connell et al. (2012) reported CPP results consistent with evidence integration in an extended noiseless detection task where participants responded to transient contrast dips in a grating stimulus. They analysed the dynamics of the CPP and its relationship with the contrast-dependent steady-state visual evoked potential (SSVEP), representing the encoding of the sensory evidence. Consistent with integration, the CPP rose at a rate predicted by the cumulative change of the SSVEP signal as the contrast dipped, and peaked at a stereotyped threshold immediately prior to the participant's response. However, since the evidence itself was increasing over time in these dips, it is difficult to definitively attribute this to integration, and the authors did not investigate if an extreme leak model could account for the CPP's dynamics.

O'Connell et al. also reported that trial-to-trial variability in the SSVEP predicted reaction time independent of the physical evidence. This may indicate that internal sensory noise fluctuations exerted a significant influence on the evidence accumulation process, making the evidence stream less reliable and more volatile. Although the task was relatively easy, this intrinsic noise may have resulted from the prolonged monitoring it required, which would be more likely to induce attentional lapses or fatigue. Indeed, changes in attentional state, which are expected to modify the level of internal noise (Maunsell & Treue, 2006; Saproo & Serences, 2014), have previously been shown to affect CPP build-up, such that signatures of better attentional engagement predict faster CPP build-up and reaction times (Kelly & O'Connell, 2013). This work provides tantalising hints that internal noise may play an influential role in determining the quality of the evidence entering the decision process which, just as with external stimulus noise, could be counteracted by evidence accumulation.

Summary

Evidence accumulation has previously been inferred from trends in behavioural data or fits with integration models, but computational modelling work has demonstrated that a more detailed approach is needed in order to exclude alternative accounts. Extreme leak architectures can mimic behavioural patterns often thought to be sufficient evidence of integration, and decision models with strong leak or short temporal filters represent viable alternatives that need to be tested. The analyses presented in this chapter test the predictions of standard integration models against those invoking substantial leak or no-integration by profiling the influence of brief evidence pulses on a number of decision-relevant behavioural and electrophysiological markers in a low-noise contrast discrimination task. A number of the studies described above that have successfully demonstrated integration without extensive modelling have also exploited a pulse paradigm (e.g. Huk & Shadlen, 2005; Kiani et al., 2008; Waskom & Kiani, 2018). This is because a simple prediction can be made about the impact of pulses depending on the presence/absence of evidence accumulation. If participants do not integrate evidence or leak evidence outside a small window, a pulse should only affect trials with reaction times shortly after the pulse. On the other hand, if participants do integrate evidence, the pulse should exert an enduring influence on trials with reaction times far exceeding the time-constants of extreme leak strategies.

Finally, as discussed in Chapter 4, analyses of the CPP were also conducted to address two important questions about its origins. First, an analysis of the latency of the CPP response to the pulses relative to a electrophysiological index of motor preparation, the lateralised readiness potential (LRP), assessed the characterisation of the CPP as representing a stage of the decision process upstream from response preparation. Second, an analysis of the CPP responses to the different kinds of pulse sought to determine if the CPP arises from a single accumulator or multiple accumulators.

Methodology

Chapter 5 and Chapter 6 report analyses of data from the same experiment. To avoid repetition, the paradigm and data processing steps will be described in detail here, and only the aspects of the data analysis that are unique to Chapter 6 will be described in its Methodology section.

Participants

12 adults participated in this study (5 female, age range: 18-39, M = 25.2 ± 6.7). However, a reliable SSVEP could not be established for one participant so they were excluded from the SSVEP analyses (n = 11). A second participant did not exhibit a reliable LRP, so they were excluded from the LRP analyses (n = 11). Both participants were retained for all other analyses (n = 12).

All participants reported normal or corrected-to-normal vision, no history of migraine or bad headaches, no history of epilepsy, and no sensitivity to flashing light. Participants provided informed written consent prior to testing and were paid a gratuity of €10 per hour of participation as compensation for their time. Participants received an additional €20 for completing all testing sessions. All procedures were approved by the Trinity College Dublin School of Psychology Ethics Committee and were in accordance with the Declaration of Helsinki.

Apparatus

The experiment was conducted in dark, sound-attenuated testing booths. Visual stimuli, generated using Psychtoolbox (Kleiner, Brainard, & Pelli, 2007) and a custom Matlab experimental script, were presented on one of two displays, either a 51 cm or a 40.5 cm gamma-corrected CRT monitor (both monitors had a 1024x768 resolution, 100 Hz frame rate). A chin rest was used to reduce head movement and ensure that the viewing distance was 60 cm. Continuous EEG data were acquired from 128 scalp electrodes using a Biosemi ActiveTwo system. Eye movements were recorded using two vertical electrooculogram (VEOG) electrodes placed above and below the left eye.

Experimental Task

Participants completed a two-alternative forced-choice contrast discrimination task, where they decided which of two orthogonally-oriented overlaid gratings was of higher contrast (see Figure 5.2). The stimuli were square-wave gratings with a spatial frequency of 1 cycle per degree of visual angle. Each grating was tilted by 45° relative to the vertical midline (right tilt at +45° and left tilt at -45°). The gratings were annular with an inner radius of 0.3° visual angle and an outer radius of 4° of visual angle, presented centrally on a grey background with the same mean luminance. The contrast levels of the two gratings were reciprocal, so if the target[1] grating was set to 60% contrast, the non-target grating would be set to 40% contrast. The differential contrast was set during a calibration phase at the beginning of the session (see Procedure section). The starting phase of the gratings was shifted a half-cycle on every trial to reduce the potential influence of adaptation. The stimulus was designed to evoke a steady-state visual evoked potential (SSVEP) to provide an independent measurement of the neural representation of the sensory evidence for each of the grating stimuli. This was achieved by separately 'frequency-tagging' the gratings by reversing their phase at 20 and 25 Hz. The frequency modulation was randomly counterbalanced between the gratings across trials. A yellow fixation point, with a radius of 0.3° of visual angle was presented in the central annulus to reduce eye movements.

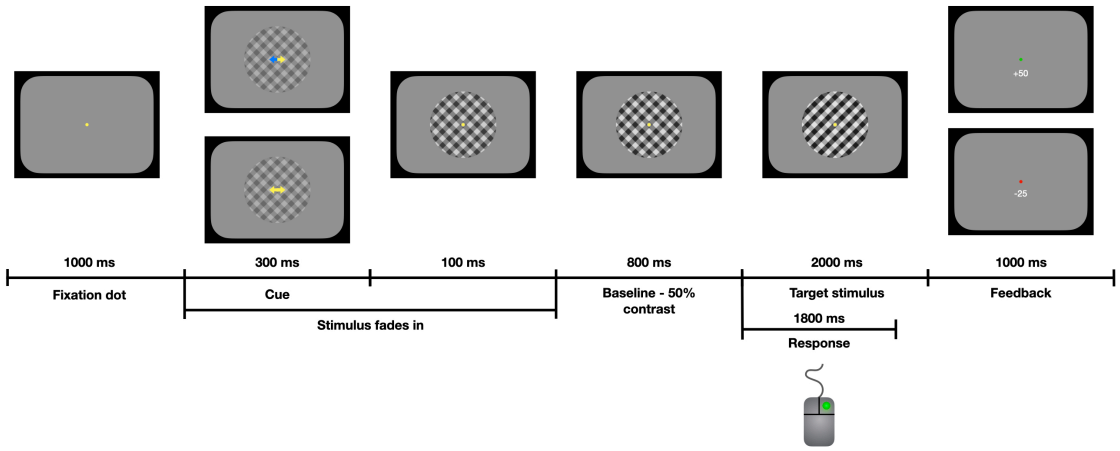


Figure 5.2. *Task schematic. Following a 1000 ms fixation phase, the stimulus gradually faded in until both gratings are presented at a baseline contrast level of 50%. The cue, which could be valid (top version), invalid (opposite of valid), or neutral (bottom version), was presented in the first 300 ms of this phase. The baseline contrast level was then maintained for 800 ms. After the baseline phase, the evidence onset, with differential contrast favouring the target stimulus. The evidence was presented for 2000 ms. Pulses, lasting 150 ms, could onset from 180 ms to 500 ms into this evidence presentation phase. The response window was 1800 ms. After the stimulus offset, the participant received feedback on their response.*

Trials began with a 1000 ms fixation period, where only the fixation point was displayed. This was followed by a 400 ms fade-in sequence where both gratings gradually increased from 0% to 50% contrast. The stimulus was

[1] Throughout this chapter, the grating that increases in contrast at evidence onset is referred to as the 'target' and the grating that decreases in contrast as the 'non-target'.

introduced gradually to avoid visual evoked potentials, which would be expected in response to a sudden change in luminance and could obscure adjustments made in response to the cue. The fade-in was followed by an 800 ms baseline period, where the gratings maintained a 50% contrast level. After this, the evidence was presented, which involved the contrast of the target grating increasing by an amount determined by a QUEST procedure (described in greater detail in the Procedure section; Watson & Pelli, 1983) and the non-target grating reducing contrast by the same amount. The target stimulus was shown for 2000 ms and the response window was 1800 ms. The stimulus presentation extended past the response window to facilitate time-frequency analyses requiring a 400 ms window. After this, a feedback screen was shown for 1000 ms. If the response was correct, the fixation point turned green and participants were shown that they had received 50 points; if the response was an error, the fixation dot turned red and participants were shown that they had lost 25 points. If the participant responded before the target onset or if they failed to respond within 1800 ms, the fixation point turned red, they were informed that they had lost 25 points, and one of two messages was displayed beneath the fixation point: "TOO LATE" or "TOO FAST! Wait for the stimulus to change!".

One of three cues were presented for 300 ms during the fade-in sequence at the beginning of each trial. Valid cues identified the tilt of the target stimulus on the subsequent trial; invalid cues indicated the tilt of the non-target stimulus; and neutral cues provided no information about the likely answer. Valid cues were presented four times as frequently as invalid cues, making them 80% predictive; invalid and neutral cues were presented on the same number of trials in each session. To maintain consistency across conditions, the cue stimulus was always a double-sided arrow. The arrow was 0.5° visual angle wide and each arrow head was 0.5° visual angle tall at its widest point. In the valid and invalid cue conditions, one side of this arrow was blue and the other was yellow. The yellow arrow indicated the probable target tilt. In the neutral cue condition, both sides of the arrow were yellow.

On no-pulse trials, the contrast difference between the correct and incorrect grating was constant across the evidence presentation. However, to aid the investigation of the processing of sensory evidence presented in different contexts, the experimental stimulus incorporated one of three types of 'pulses' of evidence on some trials: 1) reverse pulses flipped the evidence to favour the opposite grating. For example, if the contrast level of the target grating was 60% (and therefore, the contrast level of the non-target grating was 40%), during the reverse pulse, the non-target grating's contrast was increased to 60% and the target grating's contrast fell to 40% (Figure 5.3.B); 2) a gap pulse eliminated all evidence for the duration of the pulse (i.e. the contrast of both gratings was set to 50%; Figure 5.3.C); and 3) positive pulses increased the evidence for the target grating. The magnitude of this increase was designed to be equivalent to the reverse pulse, so the differential evidence for the correct grating was added to the correct grating contrast and subtracted from the incorrect grating contrast. For example, if the contrast level of the target grating was 60%, the differential evidence would be 20% (60% - 40%), therefore the positive pulse would boost the target grating contrast to 80% and reduce the non-target grating contrast to 20% (Figure 3.5.D). The positive and reverse pulses are equivalent transformations because the total contrast change during a reverse pulse in this example is 40% (60 → 40% and 40 → 60%) and the total contrast change during a positive pulse is also 40% (60 → 80% and 40 → 20%). To

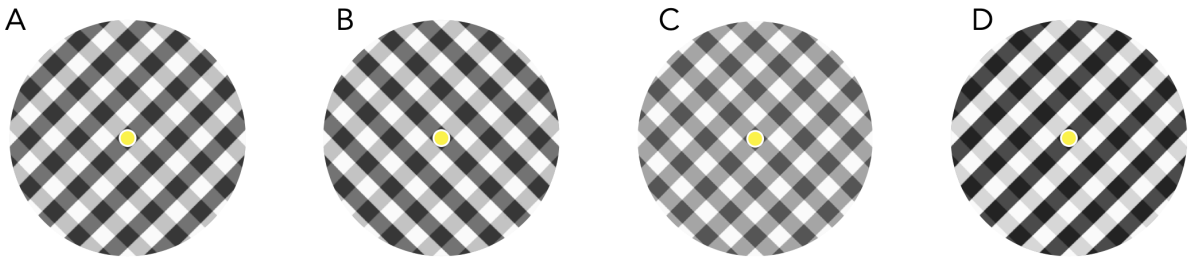


Figure 5.3. *An illustration of the different pulse types. A) No-pulse: The standard contrast difference for trials where the right-tilted grating was the target. B) Reverse Pulse: The contrast values of the two gratings are flipped to reverse the direction of the evidence. C) Gap Pulse: all evidence is eliminated as both gratings are set to 50% contrast. D) Positive Pulse: The differential evidence is doubled to boost the target evidence.*

minimise the risk of large potentials associated with sensory transients, which could obscure the dynamics of the signals being investigated, the contrast values of the gratings were always reciprocal, ensuring that the pulses were not associated with a change in the average contrast. These brief pulses were the only source of physical noise in the stimulus.

The pulses were presented for 150 ms with one of five onset times beginning at 180 ms and staggered in 80 ms increments up to 500 ms into the evidence presentation; their onset and offset were instantaneous. This short presentation was chosen to reduce the perceptual saliency of the pulse. Ideally, the pulse would impact the process of perceptual decision-making without the participant being aware that it had occurred. Indeed, most participants reported no awareness of the pulses when questioned at the end of testing. However, because this lack of awareness could not be investigated during testing, the staggering of the pulses acted as an added precaution to prevent a predictable onset being incorporated into a participant's response strategy. The specific timings were chosen based on the mean response time of a small sample collected during a pilot study. There were four pulse conditions (no-pulse, positive pulse, gap pulse, and reverse pulse) and the pulses could onset at any of five onset times, so in total there were sixteen pulse conditions. Each of these sixteen pulse conditions were equally represented in each session.

If the participant responded before the evidence onset, an exact replica of that trial was added to the end of the current block. It was also important that the participant was presented with the pulse before they responded on a pulse trial, so an extra block was included at the end of the testing session to re-present trials where responses had preceded the pulse. While it was critical to get valid responses on these trials, it was also important to avoid changing the statistics of the task since it was likely that the later pulse onset times would result in more of these premature responses than earlier pulse onsets or no-pulse trials. The extra block was devised as a method of mitigating that risk. The extra block was initially empty, however when any response occurred before the pulse onset, a set of three trials were added. One of these trials was a replica of the current trial, the remaining two trials had a different cue type and pulse-onset combination (meaning that they could have the same pulse type or pulse onset, but not both) and were randomly chosen from the remaining

alternative conditions. The replica trial was marked as immutable, but the other trials were free to vary if needed. If there was another response before pulse onset on a subsequent trial, a replica of that trial was swapped for one of the mutable trials in the extra block. If there were no more mutable trials or more than 20% of trials in the extra block had the same pulse type, a new set of 3 trials was added. If either replacement scenario occurred while the participant was currently progressing through that extra block, the replica trial was just added to the end of the block.

Procedure

Participants were asked to complete a two-alternative forced-choice task, indicating which of two overlaid gratings was presented at a greater contrast level on discrete trials. Participants responded on each trial using a mouse held in the palm of their hands with their thumbs resting on each mouse button. The experimenter demonstrated how the mouse should be held, cupped in the palms with arms outstretched on the table in front of them. To indicate the right-tilted grating had the stronger contrast, the participant pressed the right button with their right thumb and to indicate the left-tilted grating had the stronger contrast, the participant pressed the left button with their left thumb. When participants are asked to respond to incongruent stimulus-response pairings (e.g. pressing a button with their right hand in response to a left pointing arrow), they exhibit the Simon effect (Simon & Rudell, 1967), a characteristic slowing of responses in comparison to congruent stimulus-response pairings. However, in the current experiment, the right-tilted grating was always mapped to the right mouse button and the left-tilted grating was always mapped to the left mouse button. For all sessions, the participant was comfortably seated in the testing booth and placed their chin on a headrest in front of the monitor. Participants were instructed to maintain fixation on the fixation point throughout all tasks.

Calibration

Each session began with a calibration phase to individually titrate the contrast differential between target and non-target stimuli to achieve 70% accuracy. The first calibration session was composed of five stages; subsequent calibrations comprised fewer stages and were designed to allow for adjustment of these initial calibration values if participant performance changed across sessions (e.g. due to practice effects). All calibration tasks used the same overlaid gratings stimulus described in the previous section.

The participant was introduced to the task by completing a 50 trial practice block. This practice block was mostly identical to the task schematic shown in Figure 5.2, but there were no cues and the stimulus did not incorporate any pulses. Participants were told that they had 1.8 seconds to respond once the contrast changed from the baseline 50% and to familiarise them with the timing of the task, this response window would be indicated by a change in the colour of the fixation point (yellow to blue). To ensure the participant understood

the concept of the task, the first trial presented the target grating at 100% contrast and the other grating at 0% contrast (i.e. 100% evidence). The contrast differential decreased in small increments each time the participant gave correct responses on two consecutive trials and increased if they gave two consecutive error responses. Participants received feedback on each trial. The practice block was only given on the first day of testing. The timing cue used in the practice task was not included in any subsequent task.

Previous experiments conducted in the O'Connell Lab using the frequency-tagging technique indicated that the grating modulated at the higher frequency may be perceived as being presented at lower contrast than the physical stimulus (also reported by Kim et al., 2007). Two calibration tasks were designed to address this issue. In the first of these tasks, participants completed two interleaved one-up-two-down staircase procedures to estimate the contrast level that would achieve ~70% accuracy when both of the gratings were modulated by the same frequency (i.e. both 20 Hz or both 25 Hz). The differential evidence started at 100% and was adjusted by the staircase procedure according to the participant responses. The staircase ended after four reversals or after 50 trials. By comparing the contrast levels produced by these staircases, the perceptual bias could be estimated and a contrast boost could be used to compensate for the perceptual flicker effect.

The second of these tasks was designed to validate this compensatory contrast boost using a second, one-up-one-down, staircase. In this task the 25 Hz grating was always the correct answer and the 20 Hz grating was always incorrect. The staircase was designed to indicate the contrast boost to the 25 Hz grating required to achieve 50% accuracy (i.e. to make the two gratings indistinguishable), when each grating was modulated by a different frequency. The gratings were each set to 50% contrast and the 25 Hz grating was given the compensatory contrast boost estimated using the first staircase. The procedure dynamically updated this 25 Hz grating contrast boost in response to participant performance, converging on the best estimate of the perceptual bias. If there was no perceptual bias, the participant would achieve 50% accuracy when both gratings were presented at 50% contrast with no compensatory boost. The staircase ended after four reversals or 60 trials. If there was an indication of a systematic bias, this contrast boost was implemented in all subsequent tasks. However, in practice, very few participants did show a reliable bias in the first calibration session and of those that did, none continued to demonstrate this bias when these tasks were rerun in future calibration sessions.

After the flicker bias had been estimated, the participant completed a QUEST procedure (Watson & Pelli, 1983) to estimate the contrast differential (i.e. difficulty level) required to achieve 70% accuracy on the task. The procedure was initially fed the final contrast differential achieved in the practice block and then dynamically updated this estimate over 60 trials. If the previous tasks indicated a reliable perceptual bias based on flicker frequency, the stimulus incorporated the estimated compensatory boost to the 25 Hz stimulus. If the QUEST procedure estimated that a participant required the target stimulus to have a contrast level greater than 65%[2], the participant was asked to complete another practice block before repeating the QUEST procedure.

[2] A contrast level greater than 65% was deemed unacceptable because it indicated that the participant may have had insufficient practice at the task and that their performance was likely to improve with further practice.

The final phase of calibration was a simple verification that this contrast differential would result in 70% accuracy. Participants completed 30 trials with the QUEST-estimated contrast differential. If performance differed from 70% by less than 5%, the researcher adjusted the contrast differential based on the margin of error provided by the QUEST estimate and proceeded to the main task. If performance differed from 70% by more than 5%, the contrast differential was adjusted based on the QUEST margin of error and the verification block was rerun. The success of this titration procedure is demonstrated in Figure 6.7.A in Chapter 6, where accuracy in the neutral cue condition is ~70%.

Main Task

Once the calibration phase was completed, the participants started the main task. To accommodate the participants, there was a long (~1152 trials) and a short (~768 trials) version of the testing session. The trial numbers are approximate because they were subject to increases as the trial replacement mechanism was triggered. Participants completed 4-6 testing sessions. As well as cues and pulses, the main task also incorporated a points structure to motivate participants. Correct answers were rewarded with 50 points and incorrect/early/late responses were punished with -25 points. Participants were told to try to maximise their score on each block. During the main task, breaks occurred after every 48 trials, when a 20 second timeout was enforced. The break screen presented the participant with the points total scored since their last break and their average block score so far. The break screen would also present participants with both a 20-second countdown and a progress bar showing how far they were through the session. Once the timer expired, they were given the option to proceed to the next block by clicking the mouse.

Before beginning the experiment, participants were asked to maintain fixation on the fixation point at all times during the stimulus presentation. They were informed that they would receive breaks approximately every 5 minutes and that, while a 20 second break was enforced, they could take that opportunity to rest for as long as they wished and they were encouraged to take longer breaks when they felt fatigued. Participants were told that, although they were not always accurate, the cues predicted the correct answer and they should pay close attention to them, registering the direction of the cue on every trial. Participants were told to ignore the blue arrows on all trials and interpret the yellow arrow as the cue. If both arrows were yellow, this was a neutral cue and provided them with no information about the upcoming trial. The experimenter explained that they did not need to respond as fast as possible and instead, while keeping in mind the timing of the task, they should try respond when they were quite sure of their decision. Finally, the participants were not informed that there were pulses of evidence in the task. The experimenter explained that the stimulus was flickering to help with the analysis of the neural signals and that they could ignore these rapid changes. The pulses were quite difficult to perceive amongst the gratings' frequency-tagged modulations, even when one was aware that there were pulses in the task design. At the beginning of each testing session, the participant was shown the following instructions on screen: "Fixate on the central dot. Press LEFT button with LEFT hand when the LEFT-tilted pattern gets stronger. Press RIGHT button with RIGHT hand when the RIGHT-tilted pattern gets stronger."

Data Analysis

In this paradigm, pulses could onset at one of five onset times between 180 ms and 500 ms after evidence onset. To ensure that the results reflected the influence of the pulse, all analyses of the effect of pulses only included trials where the response came at least 250 ms after pulse onset. Therefore, trials with later pulse onsets necessarily had slower reaction times. However, estimating the pulse effects on behavioural or electrophysiological variables requires comparisons between trials where there was a pulse and trials where there was no pulse. To achieve some level of equivalence for this comparison, it was necessary to attempt to replicate the pulse onset reaction time restrictions in the no-pulse condition. This was achieved by dividing the no-pulse data into 'slices' by imposing equivalent reaction time restrictions on the full no-pulse dataset for each pulse onset time (see Figure 5.4). For example, for the 180 ms pulse onset time, all no-pulse trials with reaction times greater than 430 ms were included[3]. For the 260 ms onset time, all no-pulse trials with reaction times greater than 510 ms were included, and so on.

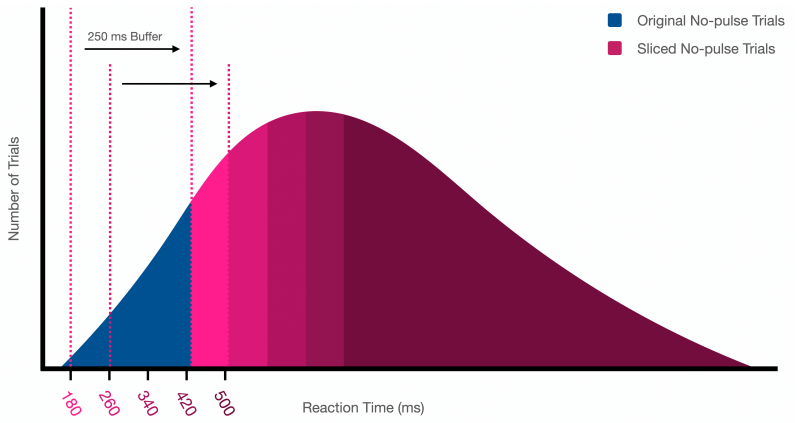


Figure 5.4. *No-pulse slicing procedure illustrated with a hypothetical reaction time distribution. To replicate the reaction time restrictions placed on the other pulse conditions, the full set of no-pulse trials was repeatedly sampled for reaction times exceeding each pulse onset time plus the 250 ms buffer. The illustration shows the portion of the full set of trials (blue) included in each of sliced onset times (darker shades of pink).*

The no-pulse trials are the intuitive control condition for analyses of the pulse effects, so the slicing procedure was used in all of the analyses of pulse effects reported in the Results section. However, since the data for each pseudo-onset time were selected from the full no-pulse dataset, this meant that a given no-pulse trial could be included in several or all slices. By design, the no-pulse condition had only 20% as many trials as the other pulse conditions with five pulse onsets. On average, the sliced trial count was 4.3 times larger following the slicing procedure. The inclusion of these replicated trials may raise concerns about falsely inflating the power of those statistical analyses using a mixed effects approach. To address any such concerns, all analyses that used

[3] Equivalent to the reaction time restriction for the earliest pulse onset (180 ms + 250 ms).

the sliced no-pulse data were repeated, removing all no-pulse trials and using the gap pulse as the control condition (see Chapter Appendix). This did not meaningfully change any of the results.

Most of the statistical analyses used mixed effects modelling. This was particularly important in this study because there was a small sample, but a large dataset for each participant. Mixed effects analyses exploit these trial numbers to reduce the risk of Type II error. There is no consensus on how to test for fixed effects using mixed effects models. However, it has been recommended that studies with small sample sizes estimate the degrees of freedom using a Satterthwaite approximation and compute an F-statistic (Kuznetsova et al., 2017; Luke, 2017). This approach was adopted for all mixed effects analyses. A random intercept was included in all models to account for the repeated-measures nature of the data, but no random variables were included in any analysis. For some analyses, the data were not amenable to a mixed effects analysis and a repeated-measures ANOVA was used instead. In these cases, where Mauchly's test indicated that the assumption of sphericity had been violated, a Greenhouse-Geisser correction was applied to the degrees of freedom. The significance of all factors included in each analysis is reported. Unless stated otherwise, all main effects were investigated with uncorrected pairwise comparisons.

Behavioural Analysis

The behavioural analyses followed a shared structure: A) confirmation of a pulse effect on the behavioural metric, and B) estimation of the duration of that pulse effect using a binned analysis (see below). Analyses attempting to profile the longevity of the pulse effect on behaviour measured reaction time relative to pulse-onset. A binary logistic mixed effects analysis was used to confirm the presence of a pulse effect in the accuracy data using post-pulse reaction time as one of the predictors. However, conducting a mixed effects analysis on the reaction time data would have necessitated predicting reaction time using the highly-correlated post-pulse reaction time. These variables are so closely related that there would be almost no variance remaining to be explained by the pulse-type factor. For this reason, a repeated-measures ANOVA was used to analyse the pulse effect on reaction time and the data were averaged across cue conditions.

Reaction time bins were used to characterise the duration of the pulse effect. To create these bins, reaction times were divided into equally sized quartiles for each pulse and pulse onset condition. These quartiles were then averaged across pulse onsets. It was important to prevent trial counts in any individual bin becoming small enough that a small number of trials could exert undue influence on the averaged data. For this reason, the data were pooled across cue conditions ensuring that every bin contained a minimum of 35 trials per participant in the binned accuracy analyses and 25 trials in the binned reaction time analyses. Since the direction of a pulse effect on reaction time was likely to reverse depending on whether the participant's response was correct or an error, only correct response trials were included in the reaction time analyses.

In addition to the binned analysis, the pulse effect on accuracy was estimated using a binary logistic mixed effects model. This was designed to validate the binned analysis of the accuracy data and provide a more

precise approximation of the delay between pulse onset and response necessary for the pulses to no longer exert an influence on choice accuracy. A linear, quadratic, and cubic model were compared using the Bayesian Information Criterion (BIC) to achieve the best fit to the data. These models were based on the mixed effects model used for the statistical analysis of the pulse effect on accuracy. Each model was identical to the previous one with two additional predictors: a post-pulse reaction time predictor raised to the next power and an interaction between this term and pulse-type (see Chapter Appendix for full model details). These models also included a random intercept to account for the repeated measures nature of the data. Changes in BIC scores were compared to critical Chi Square values for each model comparison to determine the level of model complexity that was statistically justifiable (see Table 1). The BIC comparisons indicated that the cubic model was a significant improvement on both the linear and quadratic models. The cubic model coefficients were then used to estimate the accuracy across post-pulse reaction times for all pulse conditions. The model estimate for each pulse condition was baselined by the estimate of the no-pulse condition to produce a simulation of the pulse effect on accuracy (replicating Figure 5.6.A).

Table 1. Accuracy Pulse Effect Model Fit Comparisons

Model Comparison	BIC Comparison	X^2_{change}	DF_{change}	Sig.
Quadratic vs. Linear Model	287350.435 - 287488.930	-138.4950	5	<0.001
Cubic vs. Linear Model	287787.640 - 287488.930	298.71	10	<0.001
Cubic vs. Quadratic Model	287787.640 - 287350.435	437.205	5	<0.001

EEG Analysis

Continuous EEG data were acquired from 128 electrodes using a BioSemi ActiveTwo system and digitised at 512 Hz. The data were analysed with custom scripts in Matlab using the EEGLAB toolbox (Delorme & Makeig, 2004). The EEG data were detrended and low-pass filtered below 40 Hz. Channels identified as uniquely noisy using a custom channel variance analysis were recorded for each participant and each recording session. These channels were interpolated using spherical splines. The EEG data were then re-referenced offline using the average reference. The data were segmented into a cue-locked epoch (-1700:500 ms relative to evidence onset), an evidence-locked epoch (-400:2000 ms relative to evidence onset), and a full trial epoch (-1700:2000 ms). The cue-locked epoch was baseline corrected in the interval -1400:-1200 ms relative to evidence onset; the evidence-locked and full trial epochs were baseline corrected in the interval -600:-400 ms relative to evidence onset. The full trial epoch was used in analyses predicting post-evidence signal dynamics using pre-evidence indices.

Artifact rejection was conducted separately for the cue-locked, evidence-locked, and full trial epochs. For all epochs, if the difference in activity between the VEOG channels exceeded an absolute value of 250 μV or if the voltage recorded by any scalp electrode exceeded 100 μV at any time during the epoch, that trial was excluded. After artifact rejection, a response-locked epoch (-600:400 ms relative to the response) was created from the evidence-locked data. To compensate for the effects of volume conduction across the scalp, each epoch was subjected to a Current Source Density (CSD) transformation (Kayser & Tenke, 2015). This technique is used to minimise the spatial overlap between functionally distinct EEG components (Kelly & O'Connell, 2013).

The analyses of latency and duration of the pulse effects on the CPP were compared with these metrics derived from the LRP, an index of motor preparation. As an ERP, the LRP has the advantage of sharing the same millisecond temporal resolution as the CPP, which was critical for these timing analyses. The LRP was calculated as the ERP over parietal electrodes contralateral to the response minus the ERP over parietal electrodes ipsilateral to the response. The analysis of pre-stimulus sensory encoding used an SSVEP, which is evoked over occipital electrodes by flickering the stimulus at a specific frequency. The amplitude of this oscillatory signal monotonically increases as stimulus contrast increases, so it is commonly used as an electrophysiological marker of the sensory encoding of contrast stimuli (Norcia et al., 2015; O'Connell et al., 2012; Steinemann et al., 2018).

Electrode Selection

To select CPP electrodes for each participant, a pool of central electrodes (see Chapter Appendix) were ranked according to three characteristics: 1) The mean slope of the signal over the 600 ms preceding the response; 2) The correlation between this slope and reaction time on a trial-by-trial basis; and 3) the amplitude of the signal at response (measured from -50:50 ms). The response-locked topography of the CPP was consulted to confirm that the top-ranked candidate electrodes were consistent with the known topography of the CPP. The top three suitable electrodes were selected.

For the LRP analyses, two left- and two right-hemisphere electrodes were selected for each participant separately. Electrodes were first ranked according to the mean amplitude of activity at response (higher ranks for negative amplitudes when contralateral relative to ipsilateral to the response). The top 10 ranked electrodes for left and right hemispheres were then inspected to confirm their topography was consistent with the LRP's origins over motor cortex (Ikeda & Shibasaki, 1992). It was essential that the signal was sensitive to the pulses so that the timing of its response could be estimated. For this reason, the negative-going ERPs in the top-ranked electrodes were inspected in pulse-locked plots for positive and reverse pulse trials to confirm that these electrodes detected a deflection in motor-preparation following a pulse. The response-locked activity at the chosen electrodes was then inspected to confirm that it exhibited the LRP's characteristic negative peak at

response (see Figure 5.5.D). The top two ranked electrodes that fulfilled these criteria were selected for each hemisphere.

SSVEP electrodes were chosen from an occipital candidate pool (see Chapter Appendix). Electrodes were ranked by the difference in activity associated with the target and non-target stimulus (i.e. their contrast discrimination) in the window 200:1800 ms after evidence onset. The top ranked electrodes were compared to the evidence-locked topography of SSVEP activity to confirm they were located over the occipital source. The top two suitable electrodes were chosen for each participant. There were two electrodes chosen for the SSVEP and three chosen for the CPP because the topography of the SSVEP was more focal than the CPP and, as an ERP, the CPP is a noisier signal.

Time-Frequency Analyses

The Short Time Fourier Transform (STFT) procedure was used to decompose the recording of neural activity into its time-frequency components to extract the SSVEP signals associated with the tilted gratings (20 Hz and 25 Hz). The STFT window was 400 ms, which was chosen to accommodate ten cycles of the 25 Hz SSVEP signal. The STFT was performed separately for cue-locked, evidence-locked, full-trial, and response-locked epochs. The window was moved along the length of the epoch in steps of 50 ms, providing discrete estimates of the power of neural activity at each of the frequencies of interest. Each discrete sample was mapped to the mean time-point in the 400 ms window at that position in the epoch. The SSVEP amplitude was then normalised by subtracting the amplitude of activity in the neighbouring frequency bins to isolate the stimulus-driven signal.

As mentioned in the Calibration section, all participants completed calibration procedures specifically designed to identify differences in individual perception of the grating contrast caused by the different flicker frequencies. Despite these efforts to incorporate compensatory contrast adjustments based on participant-by-participant estimates of this perceptual bias, the procedures failed to identify a stable frequency bias in most participants. Of the very few participants that did appear to exhibit such a bias, it was not observed when they were asked to complete the same calibration procedures at the beginning of their next testing session. This was interpreted as evidence that exposure to the stimuli and practice on the task had reduced any bias that was initially present. Based on this assumption, once a participant had failed to show any frequency bias, no compensatory adjustment was made to the stimuli in any subsequent session and they were not screened for a frequency bias in any subsequent session.

However, examination of the SSVEP signals after data collection revealed that this had been an error. The amplitude of the 25 Hz SSVEP was consistently greater than that of the 20 Hz signal. For this reason, in all SSVEP analyses the frequency of the target stimulus is included as a factor in the model. Its inclusion is only to account for this amplitude difference, removing that variance, so that the remaining experimental variables are isolating the appropriate effects. Therefore, its significance or non-significance is reported for each analysis, but these

results are not explored any further. There were no hypotheses based on SSVEP frequency, so it is only included as a single factor in the SSVEP analyses, but not included in any interactions with variables of interest. Examination of the behavioural data did indicate that overall, participants were more likely to choose the grating flickering at 20 Hz. However, the paradigm was carefully constructed to balance the number of trials where the target was flickering at 20/25 Hz within every condition. Therefore, it is not believed that this issue influenced any of the other effects reported in this chapter or Chapter 6. There was no difference in the number of early responses or misses depending on the flicker frequency of the target.

Timing Analyses

In this chapter, two sets of pulse-locked analyses were conducted to measure the timing of the CPP and the LRP response to pulses of evidence. The first measured the longevity of the pulse response in both signals; the second measured the relative latency of the response between the signals (discussed in Chapter 4). To optimise the signal-to-noise ratio for these timing analyses, pulse trials were pooled across cue conditions and pulse onsets, and were averaged for each participant. Only correct response trials were included since the pulse effect on the LRP would be reversed for error response trials. A 5 Hz low-pass filter was applied to the averaged LRP and CPP data and the data were baselined in the window -50:50 ms from pulse onset.

The second timing analysis focused on determining if the CPP pulse response reliably preceded the LRP pulse response, but did not necessarily aim to estimate the 'true' onset of the pulse response for either signal. For this reason, a rather conservative definition of the pulse response was used. The start of the pulse response for both signals was defined as the point halfway between A) the point of maximum positive/negative slope and B) the point of maximum positive/negative acceleration in the pulse response of the CPP and LRP respectively (see Figure 5.5). To improve the reliability of the timing estimates, the analysis was conducted on the difference between the responses to positive pulse and reverse pulse trials, targeting the point where the response to these pulses diverge. The point of maximum slope in this difference waveform was chosen as a safe indicator that the pulse response was well underway, since the steepest segment of the pulse-locked signal should occur roughly halfway to the peak of the pulse response. The point of maximum acceleration is more likely to approximate the true onset, since this is the point where the difference waveform is becoming positive/negative at the fastest rate. However, this measure was less robust to noise in the difference waveform and in some circumstances was found to underestimate the most likely onset of the pulse response. For these reasons, the average of these two points was chosen as a more reliable and fair procedure for estimating the relative timing. The procedure for detecting the pulse response was almost identical for the LRP and CPP, so for simplicity, the data processing steps will be described for the LRP and then differences in the CPP analysis will be discussed.

An algorithm was designed to detect the onset of the pulse response in the LRP and CPP based on changes in the slope and acceleration of the difference waveform for each participant. To estimate the slope, a polynomial

model was fitted to the difference waveform with a least-squares approach (using Matlab function 'polyfit'). It was important that the line faithfully reproduced the shape of the difference waveform without overfitting noise, so fits were generated for polynomials from 10 to 50 degrees. As the number of terms in a polynomial increases, the fit improves, but there is a point where the marginal benefit of another term is negligible. The residual sum of squares was plotted for each of the fits and a change point analysis was conducted (using Matlab function 'findchangepts'). This process identifies significant changes in continuous data by segmenting the data into three regions which minimises the residual sum of squared errors in each region from a line fit to that segment of the data (red line in Figure 5.5.B). In practice, this process identified the "elbow" of the residual sum of squares plot, which is the point where increasing the number of terms in the model no longer improves the fit. The model identified at the second change point was chosen as the best fit to the difference waveform and this line was plotted over the difference waveform to confirm the accuracy of the fit (purple line in Figure 5.5.A). This function was then differentiated to obtain a continuous representation of the slope of the LRP difference waveform (blue waveform in Figure 5.5.C).

The first task was to identify the deflection in the slope of the difference waveform that represented the pulse response. More negative LRP amplitudes represent greater motor preparation, so local minima were identified in the slope of the model from 50:500 ms after the pulse onset (using Matlab function 'findpeaks'; red dots in Figure 5.5.C). Starting with the negative peak closest to pulse onset, these local minima were searched for the best candidate for the pulse response. A candidate peak was adopted as the presumed pulse response if A) the slope was more negative at the candidate than the next negative peak, B) the amplitude of the difference waveform was more negative at the time of the candidate than the next negative peak, C) the most negative point in the second derivative of the model (i.e. point of maximum negative acceleration), between 50 ms after pulse onset and the next negative peak, came before the candidate. If any of these conditions were not met by a candidate, the next negative peak was tested against the same criteria. When a candidate met these criteria, it was marked as the point of maximum slope (vertical blue line in Figure 5.5.C).

Having identified the point of maximum slope in the pulse response, the next step was to identify the point of maximum acceleration. This was estimated with the second derivative of the difference waveform (yellow line in Figure 5.5.C). The second derivative was plotted from the point of maximum slope back to the previous negative peak in the slope. If the first candidate peak had been adopted as representing the pulse response, the second derivative was plotted back to 50 ms after the pulse onset. The negative peak of this second derivative plot was marked as the point of maximum acceleration in the pulse response (vertical red line in Figure 5.5.C). The start of the pulse response was then marked as the time point halfway between the point of maximum acceleration and the point of maximum slope (vertical green line in Figure 5.5.C). This process is further explained in the figure legend.

The procedure for identifying the start of the pulse response in the CPP was identical to the LRP process. The only exceptions were that A) the algorithm searched for positive slope peaks and positive acceleration; and B) there was an extra criterion for identifying the pulse response in the slope of the model: the difference

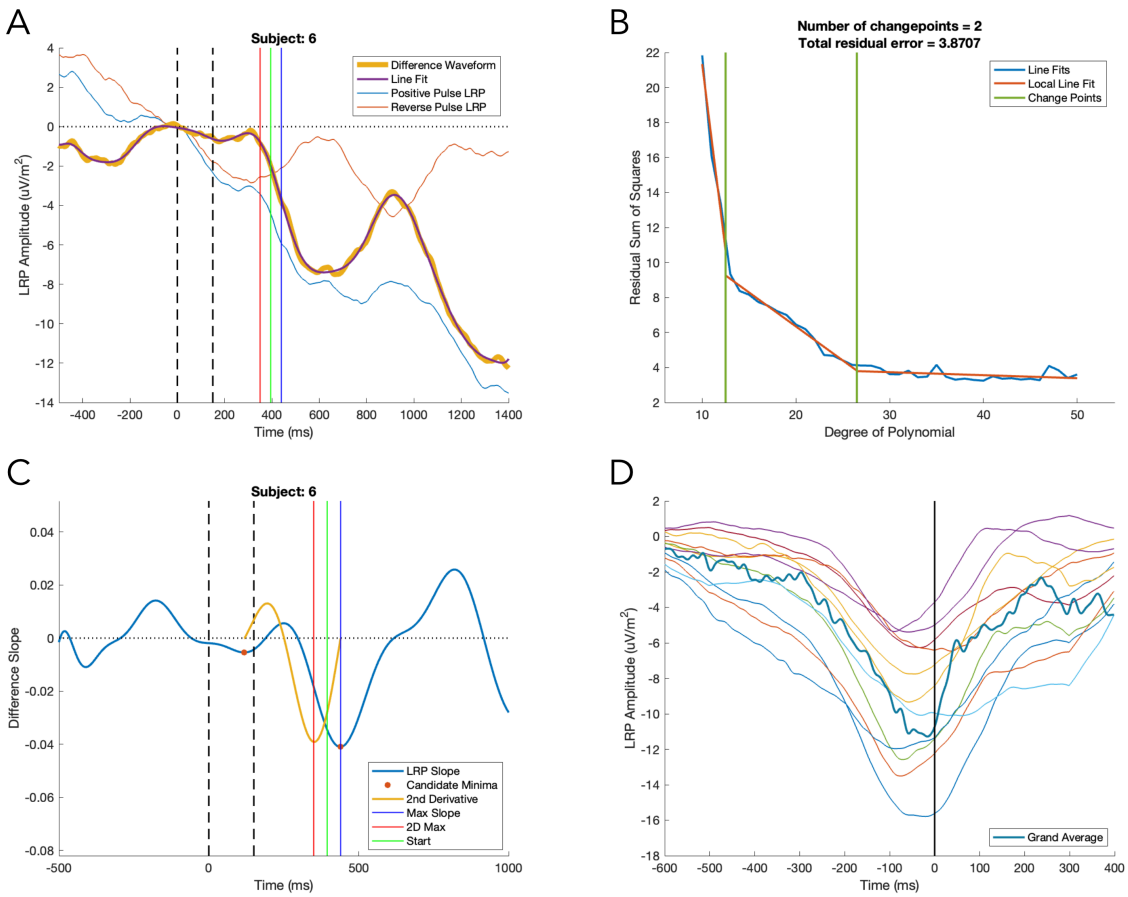


Figure 5.5. *Detecting the LRP pulse response for a single participant. A) The pulse-locked LRP for positive and reverse pulse trials is plotted relative to the correct response; the onset of the pulse responses is clearly visible in the difference waveform at approx. 400 ms. B) A model is fitted to the difference waveform by generating polynomials functions ranging from 10 to 50 degrees and plotting the residual sum of squares. A change point analysis identifies the point where increasing the number of terms in the function no longer improves the fit (second green line). This function (purple in plot A) is plotted over the difference waveform to confirm it accurately reproduces its shape. C) The chosen polynomial function is differentiated and the slope of the difference waveform is plotted. The pulse response (vertical blue line) is identified from local minima in the slope from 50 ms after pulse onset. The second derivative, representing the acceleration of the difference waveform, is plotted from the chosen negative peak in the slope to the previous negative peak or to 50 ms after pulse onset. Here, the plot of the second derivative has been amplified by 50 so it can be seen more clearly. The negative peak of the second derivative (2D max) is highlighted with a red vertical line. The start of the pulse response (vertical green line) is defined as halfway between the point of maximum negative slope and the point of maximum negative acceleration and all three markers are plotted on the original difference waveform. D) Each participant shows the characteristic LRP peak at response. Individual participant LRPs with a 5 Hz filter are shown as the thinner lines. The grand average is shown unfiltered as the thicker line.*

waveform had to be positive at the candidate time point. In other words, the CPP at the candidate peak had to show that a positive pulse led to a greater amplitude than a reverse pulse. Finally, the estimated onset of the pulse response was compared between the CPP and LRP with a paired samples t-test.

Results

The Longevity of Pulse Effects on Behaviour

To determine if the pulses of evidence had a lasting impact on participant behaviour or exhibited a pattern more consistent with an extreme leak strategy, the influence of the pulse was investigated based on reaction times relative to pulse onset (i.e. using post-pulse reaction times). Greater post-pulse reaction times entail longer delays between the pulse and the ultimate response, providing an estimate of when the pulse effect subsided. The 'pulse effect' was calculated by subtracting the average post-pulse reaction time/accuracy in the no-pulse condition from that of a particular pulse-type.

The accuracy data (see Fig 5.6.A) were analysed using a binary logistic mixed effects model. There was a main effect of Pulse-Type ($F(3,62006) = 122.521$, $p < 0.001$) and Post-Pulse Reaction Time ($F(1,62006) = 505.952$, $p < 0.001$), and a significant interaction ($F(3,62006) = 40.504$, $p < 0.001$). Accuracy was significantly greater on positive pulse trials ($p < 0.001$; $\Delta = 0.058 \pm 0.011$), and significantly reduced on gap pulse ($p < 0.001$; $\Delta = -0.024 \pm 0.09$) and reverse pulse trials ($p < 0.001$; $\Delta = -0.058 \pm 0.011$), compared to the no-pulse condition. The interaction was interpreted using the model coefficients. The magnitude of both the increase in accuracy associated with positive pulses ($\beta = -0.547$, $p < 0.001$) and the decrease in accuracy associated with reverse pulses ($\beta = 0.517$, $p < 0.001$) significantly declined as post-pulse reaction time increased. There was no significant difference in the change in accuracy in the gap pulse and no-pulse conditions across post-pulse reaction times ($\beta = 0.122$, $p = 0.189$).

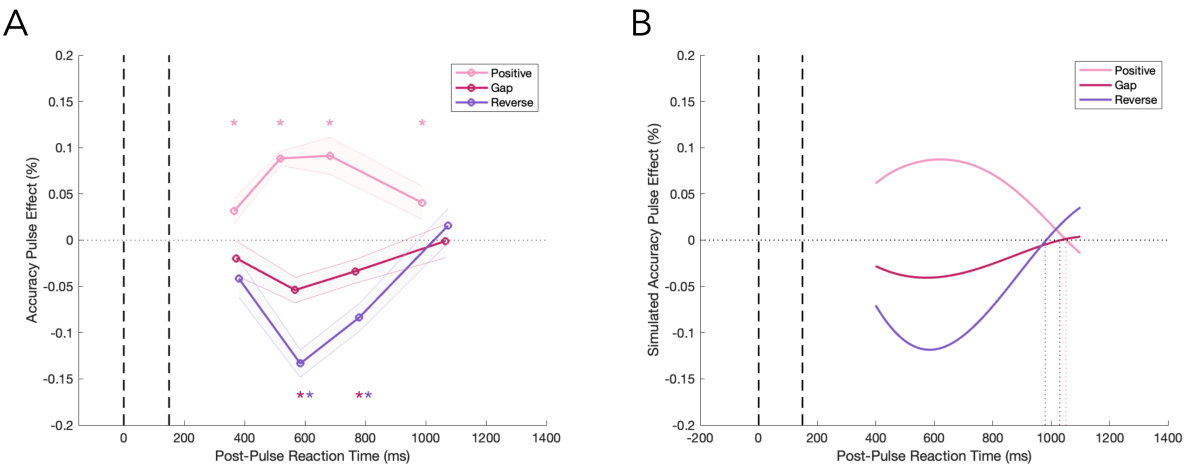


Figure 5.6. *The effect of pulses on choice accuracy across reaction times relative to pulse onset. The 'pulse effect' is simply the change in accuracy associated with the pulse, where zero represents the accuracy in the no pulse condition. The dashed lines represent pulse onset and offset. A) The binned analysis; the shaded regions represent the standard error of the mean and asterisks indicate a significant pulse effect in the highlighted reaction time bin. B) A simulation of the pulse effect on accuracy, indicating that the pulse effect fully erodes after approximately 1000 ms.*

This analysis establishes that there was a pulse effect on accuracy that lessened as the delay between pulse onset and the response increased, but it does not reveal how long after the pulse onset the pulse still made a significant contribution to accuracy. This was addressed in a follow-up analysis using paired-sample t-tests to compare the accuracy between the no-pulse condition and each of the pulse conditions for each of four reaction time bins. The results of this analysis are highlighted with asterisks in Figure 5.6. Table 2 shows the results of the binned comparisons between the no-pulse condition and each of the pulse conditions, where significant comparisons are shown in colour. The positive pulse exerted a significant effect on accuracy across all reaction time bins. For both gap and reverse pulses, there was a significant decrease in accuracy in the middle two bins.

Table 2. Paired Sample t-tests of Accuracy Pulse Effect Across Reaction Time Bins (sig. in colour)

Comparison	Mean Difference	SE	t	Sig. (2-tailed)
Positive - Bin 1	0.031	0.013	2.363	0.038
Positive - Bin 2	0.088	0.008	11.521	0.000
Positive - Bin 3	0.091	0.020	4.598	0.001
Positive - Bin 4	0.041	0.018	2.266	0.045
Gap - Bin 1	-0.020	0.019	-1.052	0.315
Gap - Bin 2	-0.054	0.014	-3.918	0.002
Gap - Bin 3	-0.034	0.013	-2.637	0.023
Gap - Bin 4	-0.001	0.018	-0.055	0.957
Reverse - Bin 1	-0.041	0.020	-2.042	0.066
Reverse - Bin 2	-0.133	0.015	-9.035	0.000
Reverse - Bin 3	-0.083	0.015	-5.591	0.000
Reverse - Bin 4	0.016	0.019	0.834	0.422

This binned analysis confirms that the pulse effect on accuracy wanes with greater intervals between the pulse onset and response, but the temporal resolution is very poor. To validate the results of the binned analysis and provide an improved approximation of this window, the pulse effect duration was estimated with a cubic binary logistic mixed effects model using post-pulse reaction time and pulse-type to predict choice accuracy (Figure 5.6.B). Consistent with the binned analysis, the model-derived estimate indicated that accuracy becomes equivalent between pulse and no-pulse conditions approximately 1000 ms after pulse onset.

Next, the post-pulse reaction time data were analysed with a repeated-measures ANOVA. The effect of a pulse on reaction time was likely to change depending on whether the participant's response on that trial was correct or was an error, so only correct responses were included in this analysis. There was a main effect of Pulse-Type ($F(1.693,18.621) = 29.375$, $p < 0.001$) and a significant Pulse-Type by Reaction Time interaction ($F(2.933, 32.264) = 19.252$, $p < 0.001$). Compared to the no-pulse condition, post-pulse reaction times were significantly faster following a positive pulse ($p < 0.001$; $\Delta = -0.048 \pm 0.018$) and significantly slower following a reverse pulse ($p = 0.002$; $\Delta = 0.027 \pm 0.015$). There was no significant difference between the gap pulse condition and the no-pulse condition ($p = 0.172$; $\Delta = 0.009 \pm 0.013$). There was also a trivial main effect of the Reaction Time Bin ($F(1.496, 16.455) = 1318.822$, $p < 0.001$) reflecting that, by definition, mean reaction times increase for later reaction time bins. The Pulse-Type by Reaction Time interaction was characterised with separate repeated measures ANOVAs for each pulse condition, using the mean post-pulse reaction time baselined by the no-pulse condition in order to isolate the pulse effect. Reaction Time Bin was the only factor in these analyses. There was a significant increase in the speeding effect of positive pulses as post-pulse reaction time increased ($F(1.301,14.309) = 13.052$, $p = 0.002$). There was also a significant increase in the slowing effect of gap pulses ($F(3,33) = 3.087$, $p = 0.041$) and reverse pulses ($F(1.695,18.642) = 8.628$, $p = 0.003$) as post-pulse reaction times increased.

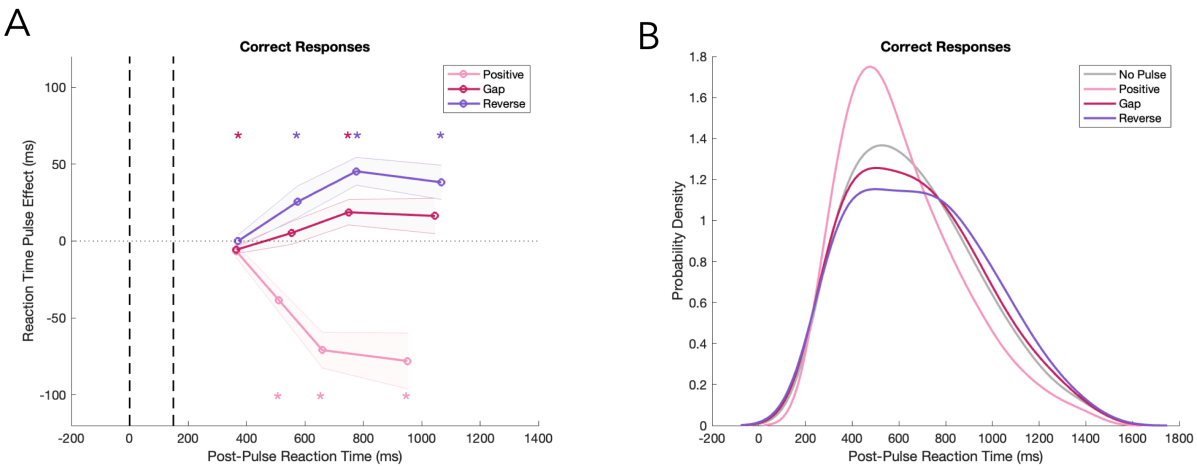


Figure 5.7. *The pulse effect on reaction time. A) The effect of pulses on reaction time across reaction time bins, relative to pulse onset. The 'pulse effect' is the change in reaction time associated with the pulse, where zero represents the reaction time in the no-pulse condition. The shaded regions represent the standard error of the mean and asterisks indicate a significant pulse effect in the highlighted reaction time bin. B) The post-pulse reaction time distribution for each pulse condition.*

A more specific indication of the emergence of this effect was provided in a follow up analysis using paired-sample t-tests to compare reaction times between the no-pulse condition and each of the pulse conditions for each reaction time bin. The results of this analysis are presented in Table 3 and highlighted with asterisks in Figure 5.7.A. Consistent with the accuracy analysis, the positive and reverse pulses seem to have clear effects on reaction time, but, unlike the effects on accuracy, they do not appear to diminish even by the final reaction

time bin (~1000 ms). The gap pulse exhibits a similar profile, but has a much smaller influence on reaction time than the other pulse types.

Table 3. Paired Sample t-tests of Reaction Time Pulse Effect Across Reaction Time Bins (sig. in colour)

Comparison	Mean Difference	SE	t	Sig. (2-tailed)
Positive - Bin 1	-0.007	0.004	-1.554	0.148
Positive - Bin 2	-0.038	0.008	-4.986	0.000
Positive - Bin 3	-0.071	0.012	-6.104	0.000
Positive - Bin 4	-0.078	0.018	-4.293	0.001
Gap - Bin 1	-0.006	0.002	-2.255	0.045
Gap - Bin 2	0.005	0.007	0.703	0.497
Gap - Bin 3	0.019	0.008	2.270	0.044
Gap - Bin 4	0.016	0.012	1.417	0.184
Reverse - Bin 1	-0.000	0.004	-0.003	0.998
Reverse - Bin 2	0.025	0.010	2.479	0.031
Reverse - Bin 3	0.045	0.009	5.031	0.000
Reverse - Bin 4	0.038	0.011	3.428	0.006

The Pulse Effect Duration Estimated from Electrophysiological Correlates of the Decision

The CPP and the LRP were used as electrophysiological indices of evidence accumulation and response preparation, respectively, to characterise the relative duration of the response to a pulse of evidence at each of these stages of the decision-making process. The LRP was calculated as the ERP contralateral to the response minus the ERP ipsilateral to the response. Since the effect of a pulse on response preparation is dependent on whether that response is correct or an error, only correct response trials were analysed.

Mixed effects analyses were conducted on the slope of each signal in the window 200:500 ms after pulse onset to confirm that there was a reliable response to the pulses (Figure 5.8). There was a main effect of Pulse-Type on the slope of the CPP ($F(3,34367.330) = 36.234$, $p < 0.001$). Compared to the no-pulse condition, there was a significant increase in the slope of the CPP after a positive pulse ($p < 0.001$) and a significant reduction in the CPP slope after a gap pulse ($p = 0.036$) or reverse pulse ($p = 0.021$). There was also a main effect of Pulse-Type on the slope of the LRP ($F(3,32676.986) = 11.737$, $p < 0.001$). Compared to the no-pulse condition, the slope of the LRP was significantly greater following a positive pulse ($p = 0.003$) and significantly reduced following a

reverse pulse (p = 0.01). There was no significant difference between the gap pulse and no-pulse conditions (p = 0.218).

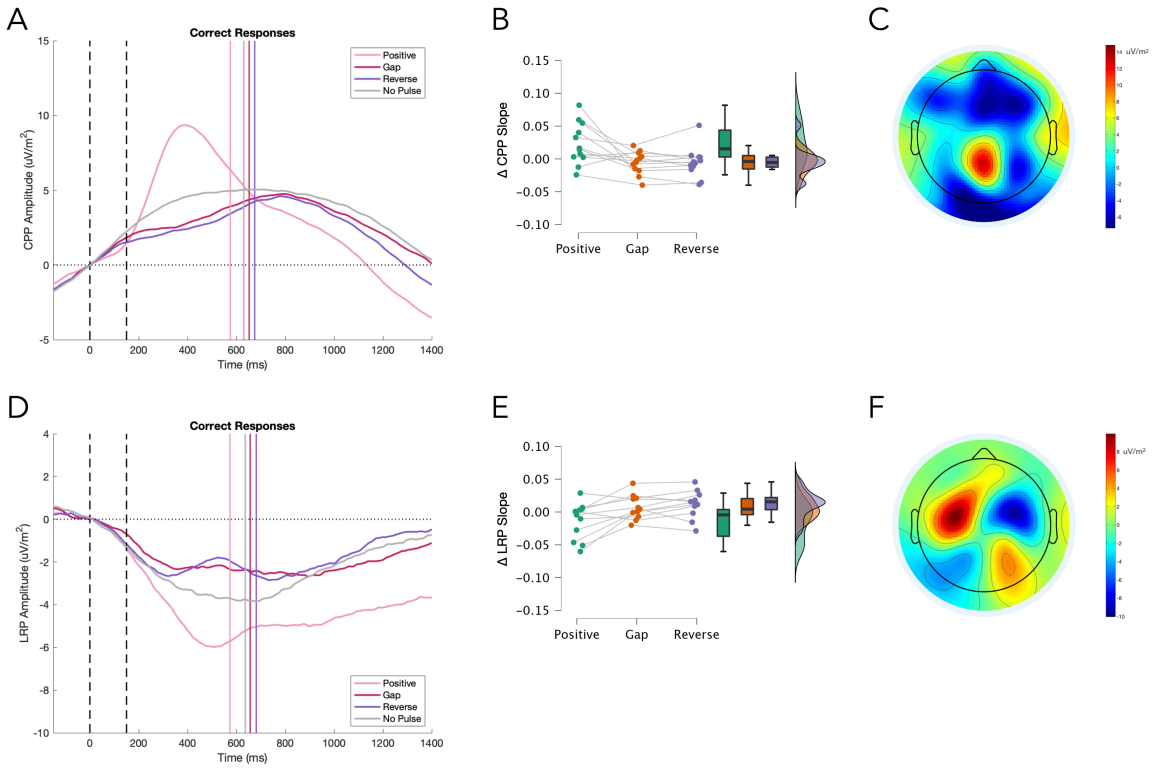


Figure 5.8. *A & D) The effect of pulses on the CPP and the LRP for correct responses. The vertical lines represent the median reaction time in the associated pulse condition. B & E) individual differences in pulse effects on the slope of the CPP and LRP, relative to the no-pulse condition C) The grand average response-locked topography, showing the CPP. F) The grand average response-locked topography for left minus right responses.*

This result confirms that the CPP and LRP are sensitive to the pulses of evidence, but it is difficult to estimate the duration of the pulse effect in the plots shown in Figure 5.8, since later time-points are more likely to include trials where the participant has already responded and the CPP and LRP are returning to baseline. To better identify the duration of pulse effects on these signals, new trial-averaged plots were generated, in which data from a given trial could only contribute to a datapoint in the grand average if the participant's response did not occur until at least 50 ms after this time point. To prevent imbalances in trial counts and signal-to-noise ratios across pulse conditions, a random subset of trials in the conditions with a larger trial count were taken to match the condition with the fewest trials for each participant. The construction of this 'response-less' waveform ended when fewer than 50 trials remained in any condition for any participant. The grand average waveforms are plotted up to this point in Figure 5.9. This ensures that the plots reflect the ongoing pre-response activity of the CPP and LRP, removing any contamination from post-response changes.

A 100 ms sliding window was used to measure the CPP amplitude along this reaction time restricted waveform in 40 ms steps for each participant and each pulse condition. These samples were compared across

participants to determine when the statistically significant difference in amplitude between positive pulse and gap and reverse pulse conditions lapsed. One-tailed[4] paired samples t-tests were performed on each of the sampled time-points after pulse onset. There were significant differences in the CPPs up to ~457 ms for the comparison of positive to gap pulse and ~500 ms for the comparison of positive to reverse pulse. For the LRP, the comparison was significant up to ~500 ms for the comparison of positive to gap pulse and up to ~574 ms for the comparison of positive to reverse pulse. This analysis was not designed to provide a precise estimate of the duration of the pulse effect on either of these electrophysiological signals. Instead, it was conducted simply to illustrate that these signals appear to be influenced by the pulse for a much longer period of time than would be expected in an extreme leak strategy.

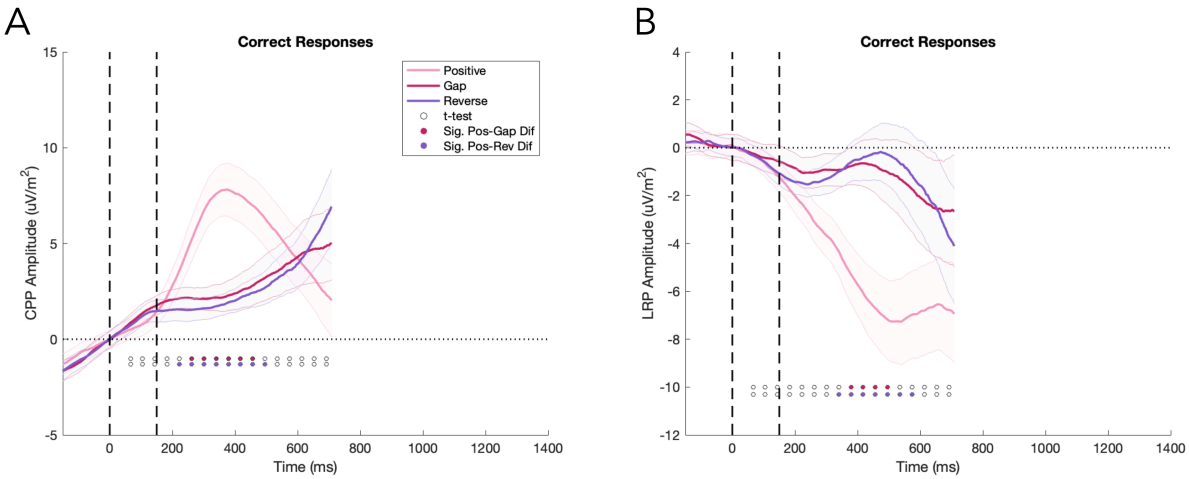


Figure 5.9. *The pulse-locked CPP (A) and LRP (B) with responses removed provides some indication of the duration of the pulse effect on these electrophysiological correlates of the decision. Each data point is the average of the trials where a response has yet to occur. The waveforms stop when fewer than 50 trials remained in any pulse condition for any participant. The shaded regions represent the standard error of the mean. A 100 ms window sampled data from these waveforms in 40 ms steps starting 50 ms after pulse onset. These samples were then compared with one-tailed paired samples t-tests to determine the duration for which the pulse effect was significant. Each comparison is represented with a black circle; the significant comparisons between positive and gap pulse conditions or positive and reverse pulse conditions are filled.*

The Relative Timing of CPP and LRP Responses To The Pulse

Having established that both the CPP and LRP are sensitive to the pulses of evidence, an analysis of the relative latency of their pulse responses was conducted using correct response trials. For each participant, the latency of the CPP and LRP responses to the pulse was estimated based on the slope and acceleration of the difference waveform, calculated by subtracting the average response to a reverse pulse from the average response to a positive pulse (Figure 5.10). The start of the pulse response was defined as the point halfway between the

[4] One-tailed tests were chosen because the direction of the pulse effects have already been established, so the cessation of the pulse effect in these comparisons could only happen in one direction.

maximum positive/negative slope and the maximum positive/negative acceleration of the difference waveform for the CPP and LRP, respectively. It was expected that this would slightly overestimate the onset of the pulse responses in both signals, since the response should be well underway at the point of maximum slope, but should provide a fair and robust measure of the relative timing. The estimated start of the pulse response was analysed with a paired samples t-test. The estimated CPP pulse response was significantly earlier than the estimated LRP pulse response ($t(1,10) = 2.624$, $p = 0.025$; $\Delta = -84 \pm 71$ ms).

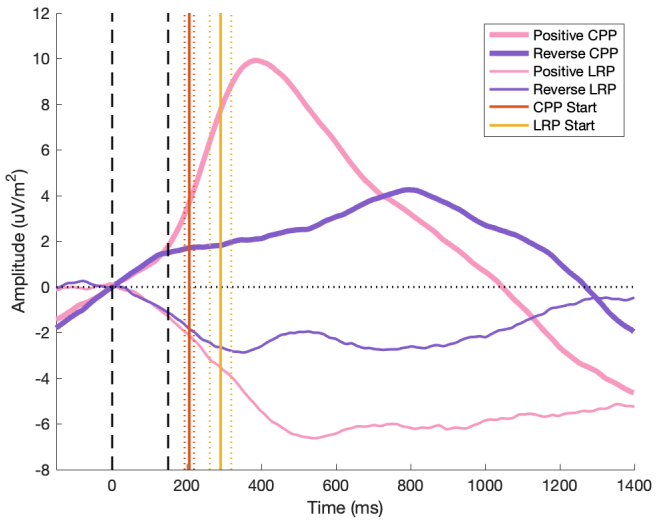


Figure 5.10. *The relative timing of CPP (positive-going) and LRP (negative-going) responses to positive and reverse pulses on correct response trials. The estimated onsets of the pulse responses are marked by vertical solid lines, the dotted lines indicate the standard error of the mean of these estimates. It is clear that the CPP response precedes the LRP response. It is also clear that these are conservative estimates and the response could be said to have started approximately 50 ms before the estimate for both signals.*

The Relevance of Internal Noise in Decision Formation

As described in the Introduction, O'Connell et al. (2012) reported that the SSVEP predicted reaction time independent of the physical evidence and this suggested there may be an important role for internal noise in decision formation. Here, the same issue was addressed by attempting to predict behaviour on a forthcoming trial using the state of the pre-evidence SSVEP on a trial-by-trial basis. Figure 5.11 shows that the SSVEP, which has a clear occipital source, rises during the fade-in sequence, plateaus during the baseline phase as contrast is held constant, and clearly separates into target and non-target at evidence onset.

The SSVEP signals associated with each grating were measured from -752:-215 ms before evidence onset. The amplitude of the SSVEP signal representing the grating that would become the non-target after evidence onset was subtracted from the amplitude of the SSVEP signal representing the upcoming target; this provided an index of the sensory representation of the stimulus referred to as the 'target marginal SSVEP'. Since the reaction

time and accuracy analyses were dependent on there being a valid response on the trial, only valid responses were included in these analyses. The role of the cue in this relationship will be investigated in Chapter 6, so it is not explored here. However, to isolate the influence of the SSVEP on decision formation, the cue must also be included in the analysis so its contribution can be segregated. The primary interest in this analysis was whether the SSVEP predicted upcoming behaviour on neutral cue trials, where the cue cannot have any influence.

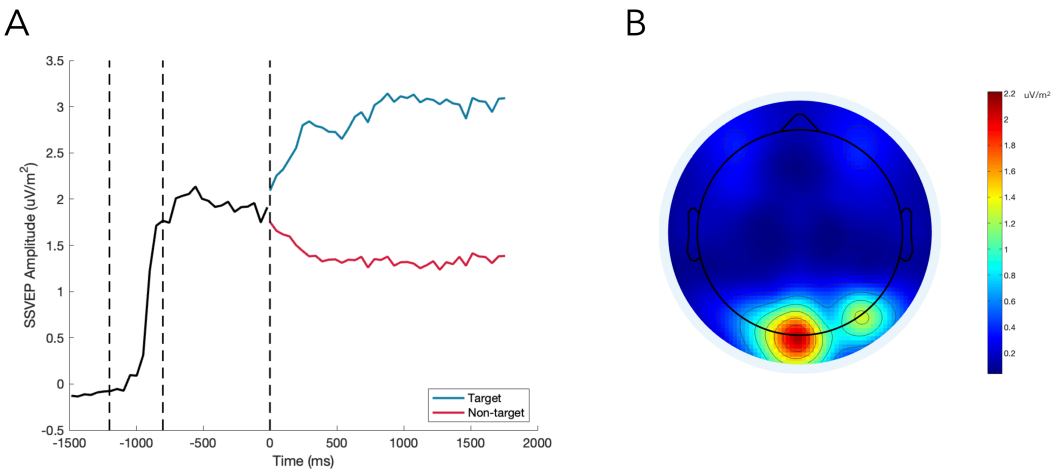


Figure 5.11. *The SSVEP tracks stimulus contrast. A) The dashed time markers represent the onset of the fade-in sequence, the onset of the baseline phase, and the onset of evidence. The mean SSVEP can be seen to rise as the stimulus fades-in. The signal then plateaus for the 800 ms baseline phase, where both gratings are presented at 50% contrast. Finally, the signal clearly discriminates the target stimulus from the non-target stimulus as evidence onsets at 0 ms. B) The topography of the SSVEP signal during evidence presentation shows strong activity over visual cortex.*

First, the data were analysed using a mixed effects analysis with SSVEP Target Frequency, Target Marginal SSVEP, Cue Condition, and a Cue-SSVEP interaction term to predict reaction time. There was a significant effect of Cue Condition ($F(2,48059.021) = 47.783$, $p < 0.001$), SSVEP Target Frequency ($F(1,48059.033) = 1099.952$, $p < 0.001$), Target Marginal SSVEP ($F(1,48059.014) = 8.372$, $p = 0.004$), and a significant interaction between Cue Condition and Target Marginal SSVEP ($F(2,48059.031) = 11.036$, $p < 0.001$). The model's Marginal SSVEP coefficient was negative and significant ($\beta = -0.002$, $p = 0.027$), suggesting that pre-evidence representation of the target leads to faster responses in the absence of a cue.

The same predictors were used to predict accuracy with a binary logistic mixed effects analysis. There was a significant effect of Cue Condition ($F(2,48069) = 2605.003$, $p < 0.001$), SSVEP Target Frequency ($F(1,48069) = 3827.074$, $p < 0.001$), and a significant interaction between Cue Condition and Target Marginal SSVEP ($F(2,48069) = 4.663$, $p = 0.009$). There was no main effect of Target Marginal SSVEP ($F(1,48069) = 0.15$, $p = 0.698$). The marginal SSVEP coefficient was negative and significant ($\beta = -0.018$, $p = 0.016$), suggesting that greater pre-evidence representation of the target leads to less accurate responses in the absence of a cue.

A final analysis was conducted to determine if the representation of the stimuli with equal contrast before evidence onset was related to the representation of the stimuli after evidence onset. The same model was used to predict the target marginal SSVEP in the window 684:977 ms during evidence presentation. This window was chosen to come after the last pulse offset, but not to push too far into the evidence-locked epoch when responses would have been made on most trials. All trials were included in the final analysis, as a response was not necessary to have an SSVEP representation of the stimulus. There was a significant effect of Pre-evidence Marginal SSVEP ($F(1,40109.038) = 305.592$, $p < 0.001$), Cue Condition ($F(2,40109.086) = 5.733$, $p = 0.003$), and SSVEP Target Frequency ($F(1,40109.151) = 1269.625$, $p < 0.001$), but no interaction between the Pre-evidence Marginal SSVEP and Cue Condition ($F(2,40109.174) = 0.739$, $p = 0.478$). The analysis was repeated without the interaction term to characterise the main effect of Pre-evidence Marginal SSVEP. Target representation during evidence presentation was significantly improved as the pre-evidence marginal SSVEP increased ($\beta = 0.134$, $p < 0.001$).

Evidence Representation in the CPP

As discussed in Chapter 4, little is known about how the neural generators of the CPP represent evidence. The CPP may be the product of global summation across multiple accumulators representing the state of evidence for each decision alternative. In this case, pulses of evidence in either direction would be expected to increase CPP build-up because any evidence would increase activity in one of its constituent accumulators. Alternatively, the combined activity of accumulators may come to represent the same quantity as the decision variable in the Drift Diffusion Model (i.e. relative evidence favouring a dominant perceptual hypothesis). This could be achieved by mutually-inhibiting pools of neurons, representing the accumulated evidence for each alternative, or the relative evidence may be expressed at the level of individual neurons; this remains an open question (Forstmann et al., 2016). Regardless of the specific neural implementation, in this account, pulses of evidence in the direction of the current hypothesis would increase CPP build-up, but this bipolar mapping of the decision alternatives would produce transient dips in the slope and amplitude of the CPP when pulses of counter-evidence were presented. To adjudicate between these accounts, the response of the CPP to pulses of evidence was analysed for correct and error response trials separately, with particular interest in the effect of reverse pulses on correct response trials and positive pulses on error response trials (i.e. the effect of counter-evidence).

A mixed effects analysis using pulse-type as the only predictor was conducted on the CPP amplitude in the pulse-locked window 250:550 ms, baseline-corrected to the interval -50:50 ms prior to pulse onset. On correct response trials, there was a main effect of Pulse-Type ($F(3,34380.085) = 63.037$, $p < 0.001$). Relative to the no-pulse condition, there was a significant decrease in CPP amplitude following a gap pulse ($p < 0.001$) or reverse pulse ($p< 0.001$), and a significant increase in CPP amplitude following a positive pulse ($p < 0.001$). There was no main effect of Pulse-Type on error response trials ($F(3,10803.321) = 1.441$, $p = 0.229$).

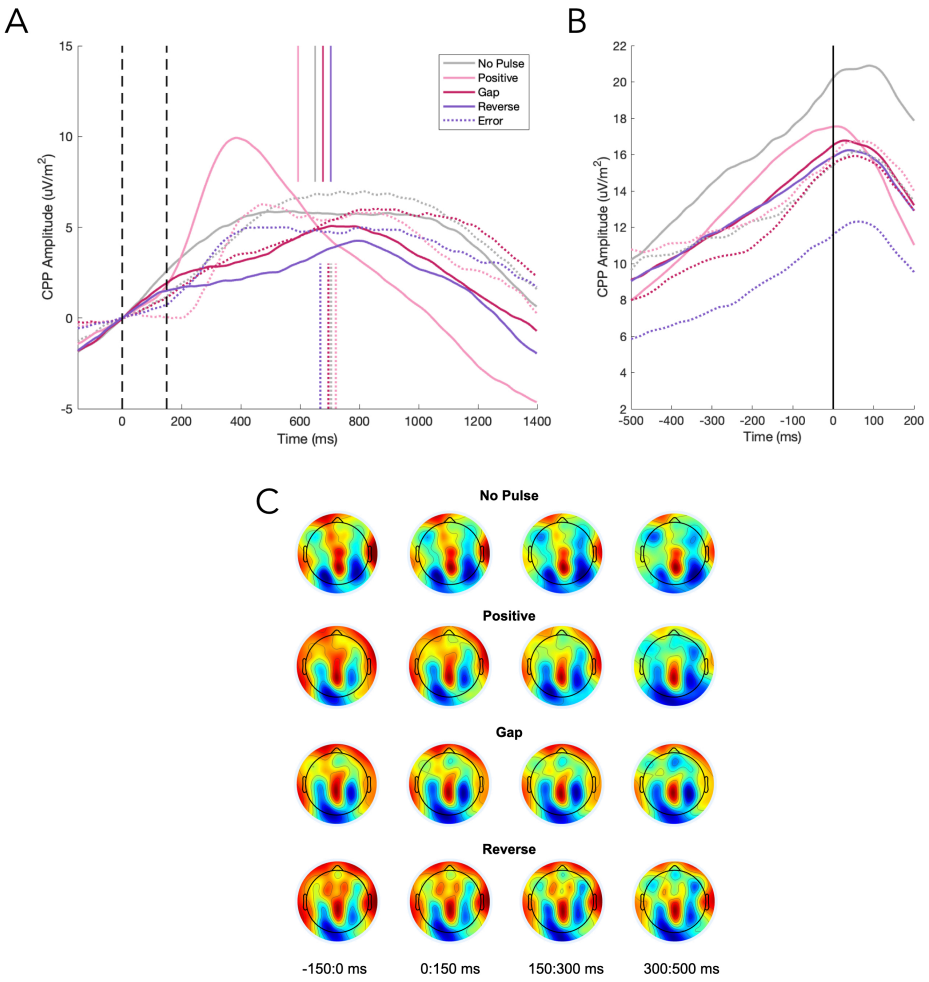


Figure 5.12. *A) The pulse-locked CPP response to pulses on correct and error trials. B) The response-locked CPP for the same set of conditions. C) Grand average topographies for each pulse-type in four windows following pulse onset. There is no indication that the CPP response to the pulse is contaminated by a visual evoked potential over occipital electrodes.*

This analysis was repeated on the CPP slope measured from 150:350 ms. On correct response trials, there was a main effect of Pulse-Type ($F(3,34381.989) = 223.945$, $p < 0.001$). Relative to the no-pulse condition, there was a significant decrease in CPP slope following a gap pulse ($p < 0.001$) or reverse pulse ($p < 0.001$), and a significant increase in CPP slope following a positive pulse ($p < 0.001$). There was also a main effect of Pulse-Type on error response trials ($F(3,10802.607) = 5.602$, $p < 0.001$). Compared to the no-pulse condition, there was a significant increase in CPP slope following a positive ($p = 0.005$) or reverse pulse ($p = 0.025$), but no significant difference between the no-pulse and gap pulse conditions ($p = 0.59$).

While these analyses describe the direction of pulse effects, the changes in amplitude and slope detected in these analyses were relative changes. That is, they indicate that the CPP amplitude and slope was reduced following a gap or reverse pulse compared to if there had not been a pulse, but not that it was reduced in an absolute sense (as can be seen in Figure 5.12.A). Since the accounts being considered differ in predictions about the absolute CPP amplitude and slope, additional analyses were included to characterise the absolute change in the CPP amplitude and slope in response to a pulse of counter-evidence. In each case, mixed effects

analyses with no predictor variables were conducted on the CPP amplitude or slope in a particular condition to determine if the intercept value (representing the post-pulse change in that parameter) differed significantly from zero.

Although the CPP amplitude exhibited a relative decrease, there was nevertheless a significant absolute increase following a gap pulse ($F(1,10.056) = 15.944$, $\beta = 2.79$, $p = 0.003$) and a reverse pulse ($F(1,8.993) = 20.522$, $\beta = 2.55$, $p = 0.001$) on correct response trials. The absence of an absolute dip following counter-evidence is more consistent with the account proposing multiple simultaneous accumulators. In addition, the slope of the reverse pulse CPP on correct response trials did not significantly differ from zero ($F(1,9.553) = 0.198$, $p = 0.666$) and the model estimate of the slope in this condition was positive ($\beta = 0.003$). The slope of the positive pulse CPP on error response trials was significantly positive ($F(1,9.518) = 11.356$, $\beta = 0.049$ $p = 0.008$). The observation that the positive pulse led to a significant increase, not a decline, in CPP slope on error response trials also favours the multiple accumulator hypothesis. To ensure that these effects were not attributable to contamination of the CPP by a visual evoked potential (VEP) elicited by the pulse stimulus, the topographies of each pulse-type were inspected in four windows across the pulse-locked epoch (Figure 5.12.C). There is no evidence of a pulse-evoked VEP in these topographies.

If two accumulators were active simultaneously, on correct trials, a reverse pulse should lead to greater CPP activity than a gap pulse; and on error trials, a positive pulse should lead to greater CPP activity than a gap pulse. This is because in both instances, the accumulator for the incorrect alternative is being provided with evidence that is absent in the gap pulse condition. However, pairwise comparisons from the previous mixed effects analyses of correct response trials showed that there was no significant difference between gap and reverse pulse conditions on the CPP amplitude ($p = 0.62$) or slope ($p = 0.75$). As mentioned previously, there was no main effect of pulse-type on CPP amplitude on error response trials. Pairwise comparisons indicated that there was a significant difference on error response trials in the slope of the CPP between positive and gap pulse conditions ($p < 0.001$). These results do not offer much support for the multiple accumulator hypothesis.

Discussion

Rigorous research efforts have established a role for evidence accumulation in the formation of perceptual choices based on stimuli that are heavily corrupted by physical noise, such as random dot motion, but relatively little attention has been paid to testing its role in other common types of decisions in which stimuli have low or no noise. Here, an empirical assessment of the latter assumption was devised. Participants completed a low-noise contrast discrimination task in which evidence was maintained at a constant level throughout each trial aside from a brief (150 ms) pulse of evidence that was designed to perturb the decision process mid-flight. These pulses exerted a clear influence on a number of behavioural and neurophysiological signatures of the decision-making process for durations far exceeding those implied by extreme leak accounts. Pulses influenced both accuracy and reaction time up to ~1000 ms after their onset and influenced the build-up of the CPP and LRP for at least 500 ms.

Several mathematical modelling studies highlight how difficult it is to definitively adjudicate between integration and extreme leak accounts based on behavioural analyses alone (Ditterich, 2006; Stine et al., 2020), but access to neural signatures of decision formation provides a powerful means of empirically testing these accounts through careful paradigm design. This experiment was partly inspired by the work of Huk and Shadlen (2005) who demonstrated that brief motion pulses had long lasting effects on behavioural (~900 ms) and neural (~575 ms) indices of decision formation in monkeys. The present findings accord well with those of Huk and Shadlen with comparable pulse effect durations on behaviour and decision signals, confirming that their results extend to humans and to perceptual choices about stimuli with little physical noise.

Using a similar pulse paradigm, Carland et al. (2016) reported that their behavioural data were better accounted for by an Urgency Gating Model than a Drift Diffusion Model. Their Urgency Gating Model included 300 ms of non-decision time, so the model actually did predict that a pulse effect on reaction time may still be evident ~550 ms after its onset, but this effect should only emerge 300 ms after pulse onset. The reaction time effects in the current dataset outlast this range, emerging in the first bin with a mean reaction time of ~400 ms and persisting to the last bin with a mean reaction time of ~1000 ms. However, this does illustrate an issue with comparing integration and extreme leak models based on reaction time. Even an Extrema Detection model may exhibit surprisingly prolonged reaction time effects. For example, if a gap pulse, which eliminates all evidence, prevents an Extrema Detection model from responding for 150 ms, the entire reaction time distribution will shift 150 ms to the right[5]. In comparison to the Extrema Detection reaction time distribution for no-pulse trials, this may give the illusion of an enduring pulse effect.

[5] Assuming all evidence samples have the same probability of crossing the bound and triggering a response.

It is for this reason that neurophysiological metrics are critical in separating these accounts. In the current dataset pulses were found to exert a sustained influence on the CPP and LRP that persisted up to 500 ms after pulse offset. Importantly, the CPP and LRP represent stages of processing that precede the motor execution component of non-decision time, so these estimates are twice as long as would be expected if sensory data were only retained in the decision process for 100-250 ms.

The present study did not uncover any definitive evidence of leakage in the evidence accumulation process. Although the magnitude of the pulse effect on accuracy can be seen to decline as reaction times extend further away from pulse onset, the current analyses are not equipped to provide any explanation for that observation. It is conceivable that this is the result of leak, but this could equally be a product of noise in the decision process on trials with particularly slow reaction times. One of the goals of this chapter was to test for a signature of leak in the CPP by introducing gaps in the physical evidence. If the CPP represents the cumulative evidence in favour of the current perceptual hypothesis, then a transient removal of the evidence would be expected to cause the signal to briefly return toward baseline as older evidence samples leak out of the decision process. No such dip in the CPP was observed here. Ultimately, modelling the present data with and without leak would be important in interpreting this result and its implications for the potential role of leak in this task.

Why would participants integrate evidence on this task when there was minimal physical noise? One explanation is provided by the observation that the state of sensory encoding before any physical evidence is presented predicts several features of the subsequent decision process on a trial-by-trial basis. Specifically, the marginal representation of the forthcoming target grating in the pre-evidence SSVEP predicted reaction time, accuracy, and the quality of the representation of the sensory evidence on the upcoming trial. Since the contrast levels of the stimuli were equal, and there is little physical noise, this result suggests there may be an important role for internal noise in decision formation in this task. There is no reason to believe that the brain discriminates between internal and external sources of noise in its extraction of the sensory evidence, so the apparent influence of internal noise provides a reason for participants to integrate evidence, even in the absence of persistent physical noise. When perceptual tasks are easy, internal noise should be drowned out by the strong sensory evidence, but when perceptual decisions are made more difficult (as they were in this study), the weak stream of sensory evidence can be polluted by internal noise and evidence integration may be an effective strategy to filter the signal from that noise.

The CPP Represents The Decision State Upstream of Motor Preparation

The CPP has been characterised as an electrophysiological signature of an abstract decision process that accumulates evidence independent of the sensory or response modality (O'Connell & Kelly, 2021). The question of whether the CPP process intermediates between sensory encoding and response preparation has only been tested once before, with Kelly and O'Connell (2013) reporting that the CPP's evidence-dependent build-up preceded that of the LRP by 150 ms. The present paradigm is perfectly tailored to replicate this

analysis, timing the CPP and LRP responses to brief pulses of evidence. The CPP pulse response emerged significantly earlier (~84 ms) than that of the LRP. This result serves to reinforce the previous observation that the CPP offers insight into a representation of the abstract decision state upstream of motor preparation. Kelly and O'Connell (2015) point out that this feature separates the CPP from other neurophysiological indices of the decision process that are tethered to some form to response preparation. They suggest that a domain-general decision process, divorced from effector-specific preparation, may facilitate the behavioural flexibility that characterises human interactions with their environment. The rapid emergence of the pulse effect in the LRP following the CPP is consistent with a decision architecture that balances this flexibility with continuous information flow to a distributed response-preparation network, enabling the execution of rapid responses and the countermanding of those responses where needed (Kelly & O'Connell, 2015).

The CPP's Representation of Evidence

One of the questions that remain about the neural basis of the CPP is, what exactly does the amplitude of the CPP reflect? Two hypotheses were considered in the final analysis. The first account suggests that the CPP is the product of multiple neural accumulators, representing the state of the evidence for each decision alternative. This hypothesis suggests that in this two-alternative task, pulses of evidence should increase the amplitude of the CPP independent of the direction of the pulse. An alternative possibility is that the CPP is the product of a single accumulator, reflecting the relative evidence in favour of the current perceptual hypothesis. In this case, pulses of evidence in the direction of the current perceptual hypothesis should increase the CPP amplitude and pulses of counter-evidence should drive a negative deflection in the CPP.

Although the CPP's amplitude was reduced relative to the no-pulse condition, in absolute terms, it maintained a positive trajectory following both gap and reverse pulses. These results appear to be consistent with the hypothesis that the CPP represents the summed activity of multiple accumulators. Yet, other aspects of the pulse responses are not readily explicable in terms of the multiple accumulator account, including the finding that there was no difference in the effect of gap and reverse pulses on CPP build-up on correct response trials and that there was no effect of pulse at all on CPP amplitude for error response trials. Indeed, the results may be compatible with a single-accumulator account that exhibits muted responses to counter-evidence.

Others have reported a negative CPP response to counter-evidence or gaps in the evidence. O'Connell et al. (2012) recorded the CPP while participants were performing a continuous target detection task with a contrast stimulus. On some trials, the slow target sequence (a 1.6 s drop in contrast) was interrupted by a return toward baseline for 450 ms. Their plots show that this gap in the evidence produced a transient dip in the CPP response, consistent with the hypothesis that the CPP represents the relative accumulated evidence. However, this was a gradual change in contrast over a period three times as long as the pulses in the current study. Additionally, they used a one-choice detection task, so this activity may represent a single accumulator and, hence, might not be directly comparable with the current dataset. Afacan-Seref et al. (2018) asked participants

to complete a simple colour discrimination task with a very short response deadline (325 ms). Similar to O'Connell and colleagues, they found that the CPP dips when participants' LRPs indicate that they initially prepared to respond to a high-value target, but ultimately responded to the correct lower-value target. Again, it is difficult to draw comparisons with the task used in this study because this may be particular to very rapid decisions and the conflict in their study was between value-based accumulation and subsequent evidence accumulation, rather than conflicting stimulus evidence.

Taken together, the present analyses cannot be said to offer any definitive evidence for either hypothesis regarding the neuronal origins of the CPP. The lack of negative CPP deflection following pulses of counter evidence is an interesting observation, but future work should explore the degree to which simulations of different possible neural instantiations of the CPP reproduce this pattern of results.

References

Afacan-Seref, K., Steinemann, N. A., Blangero, A., & Kelly, S. P. (2018). Dynamic Interplay of Value and Sensory Information in High-Speed Decision Making. *Current Biology*, *28*, 795-802.e6.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *The Journal of Neuroscience*, *12*, 4745–4765.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience*, *10*, 1157–1169.

Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science*, *340*, 95–98.

Carland, M. A., Marcos, E., Thura, D., & Cisek, P. (2016). Evidence against perfect integration of sensory information during perceptual decision making. *Journal of Neurophysiology*, *115*, 915–930.

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: the urgency-gating model. *The Journal of Neuroscience*, *29*, 11560–11571.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

de Lafuente, V., Jazayeri, M., & Shadlen, M. N. (2015). Representation of accumulating evidence for a decision in two parietal areas. *The Journal of Neuroscience*, *35*, 4306–4318.

Ditterich, J. (2006). Stochastic models of decisions about motion direction: behavior and physiology. *Neural Networks*, *19*, 981–1012.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *The Journal of Neuroscience*, *32*, 3612–3628.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. *Annual Review of Psychology*, *67*, 641–666.

Ghose, G. M. (2006). Strategies optimize the detection of motion transients. *Journal of Vision*, *6*, 429–440.

Glickman, M., & Usher, M. (2019). Integration to boundary in decisions between numerical sequences. *Cognition*, *193*, 104022.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. Annual Review of Neuroscience, 30, 535–574.

Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. Nature, 404, 390–394.

Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *The Journal of Neuroscience*, *25*, 10420–10436.

Ikeda, A., & Shibasaki, H. (1992). Invasive recording of movement-related cortical potentials in humans. *Journal of Clinical Neurophysiology*, *9*, 509–520.

Kayser, J., & Tenke, C. E. (2015). On the benefits of using surface Laplacian (current source density) methodology in electrophysiology. *International Journal of Psychophysiology*, *97*, 171–173.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *The Journal of Neuroscience*, *33*, 19434–19441.

Kelly, S. P., & O'Connell, R. G. (2015). The neural processes underlying perceptual decision making in humans: recent progress and future directions. *Journal of Physiology, Paris*, *109*, 27–37.

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *The Journal of Neuroscience*, *28*, 3017–3029.

Kim, Y. J., Grabowecky, M., Paller, K. A., Muthu, K., & Suzuki, S. (2007). Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nature Neuroscience*, *10*, 117–125.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, 36(14), 1-16.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmertest package: tests in linear mixed effects models. Journal of Statistical Software, 82.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494–1502.

Manwani, A., & Koch, C. (2001). Detecting and estimating signals over noisy and unreliable synapses: information-theoretic analysis. *Neural Computation*, *13*, 1–33.

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*, 317–322.

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex*, *13*, 1257–1269.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*, 4.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*, 1729–1735.

O'Connell, R. G., & Kelly, S. P. (2021). Neurophysiology of Human Perceptual Decision-Making. *Annual Review of Neuroscience*, *44*, 495–516.

Ossmy, O., Moran, R., Pfeffer, T., Tsetsos, K., Usher, M., & Donner, T. H. (2013). The timescale of perceptual evidence integration can be adapted to the environment. *Current Biology*, *23*, 981–986.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. *Annual Review of Neuroscience*, *21*, 227–277.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: current issues and history. *Trends in Cognitive Sciences*, *20*, 260–281.

Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: the temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General*, *139*, 70–94.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, *21*, 409–418.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of Neuroscience*, *22*, 9475–9489.

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America*, *61*, 1176–1186.

Saproo, S., & Serences, J. T. (2014). Attention improves transfer of motion information between V1 and MT. *The Journal of Neuroscience*, *34*, 3586–3596.

Steinemann, N. A., O'Connell, R. G., & Kelly, S. P. (2018). Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nature Communications*, *9*, 3627.

Stine, G. M., Zylberberg, A., Ditterich, J., & Shadlen, M. N. (2020). Differentiating between integration and non-integration strategies in perceptual decision making. ELife, 9.

Summerfield, C., & Blangero, A. (2017). Perceptual Decision-Making. In *Decision Neuroscience* (pp. 149–162). Elsevier.

Thura, D., Beauregard-Racine, J., Fradet, C.-W., & Cisek, P. (2012). Decision making by urgency gating: theory and experimental support. *Journal of Neurophysiology*, *108*, 2912–2930.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.

Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, *19*, 326–339.

Waskom, M. L., & Kiani, R. (2018). Decision Making through Integration of Sensory Evidence at Prolonged Timescales. *Current Biology*, *28*, 3850-3856.e9.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*, 113–120.

Watson, A. B. (1979). Probability summation over time. *Vision Research*, *19*, 515–522.

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, *370*, 140–143.

Appendix 5.1 - Estimating The Longevity of The Pulse Effect on Accuracy

A binary logistic mixed effects model was used to simulate the duration of the pulse effect on choice accuracy. The simulation model was selected based on a comparison of linear, quadratic, and cubic versions of the model used in the pulse accuracy analysis. Table 2 describes the structure of each of these candidate models. The novel additions to each iteration are highlighted and predictors retained from the previous model are shown in grey.

Table 4. Pulse Effect Models

Model	Predictors
Linear Model	Pulse-type
DF = 12	Post-Pulse Reaction Time
BIC = 287488.930	Pulse-type * Post-Pulse Reaction Time
Quadratic Model	Pulse-type
DF = 17	Post-Pulse Reaction Time
BIC = 287350.435	Pulse-type * Post-Pulse Reaction Time
	Post-Pulse Reaction Time * Post-Pulse Reaction Time
	Pulse-type * Post-Pulse Reaction Time * Post-Pulse Reaction Time
Cubic Model	Pulse-type
DF = 22	Post-Pulse Reaction Time
BIC = 287787.640	Pulse-type * Post-Pulse Reaction Time
	Post-Pulse Reaction Time * Post-Pulse Reaction Time
	Pulse-type * Post-Pulse Reaction Time * Post-Pulse Reaction Time
	Post-Pulse Reaction Time * Post-Pulse Reaction Time * Post-Pulse Reaction Time
	Pulse-type * Post-Pulse Reaction Time * Post-Pulse Reaction Time * Post-Pulse Reaction Time

Appendix 5.2 - Using The Gap Pulse As A Control Condition

To assuage concerns about the no pulse slicing approach adopted in the analyses reported in this chapter, the analyses were repeated, excluding the no pulse trials and using the gap pulse as the control condition instead. As shown below, this did not meaningfully change any of the results; any differences are highlighted in red.

1. **Binary Logistic Mixed Effects Analysis of Pulse Effect on Accuracy**

   There was a main effect of pulse-type ($F(2,46960) = 181.664$, $p < 0.001$), post-pulse reaction time ($F(1,46960) = 370.232$, $p < 0.001$), and a significant interaction ($F(2,46960) = 60.292$, $p < 0.001$). Contrasts showed that accuracy was significantly greater on positive pulse trials ($t(56) = 13.562$, $p < 0.001$; 95% CI = 0.083± 0.013) and significantly reduced on reverse pulse trials ($t(2453) = -6.568$, $p < 0.001$; 95% CI = -0.34 ± 0.010), compared to the gap pulse condition. The interaction was interpreted using the model coefficients. Compared to the gap pulse condition, the magnitude of both the increase in accuracy associated with positive pulses ($\beta = -0.673$, $p < 0.001$) and the decrease in accuracy associated with reverse pulses ($\beta = 0.395$, $p < 0.001$) significantly declined as post-pulse reaction time increased.

2. **Binned Analysis of Pulse Effect on Accuracy Using Paired Sample t-tests**

   In the original analysis, the reverse pulse did not have a significant effect on accuracy in the first reaction time bin.

Table 5. Paired Sample t-tests of Accuracy Pulse Effect Across Reaction Time Bins

Comparison	Mean Difference	SE	t	Sig. (2-tailed)
Positive - Bin 1	0.051	0.013	3.849	0.003
Positive - Bin 2	0.142	0.014	9.949	0.000
Positive - Bin 3	0.125	0.017	7.276	0.000
Positive - Bin 4	0.042	0.012	3.607	0.004
Reverse - Bin 1	-0.022	0.009	-2.422	0.034
Reverse - Bin 2	-0.079	0.008	-10.446	0.000
Reverse - Bin 3	-0.050	0.013	-3.792	0.003
Reverse - Bin 4	0.017	0.011	1.526	0.155

3. Repeated Measures ANOVA on Reaction Time Data

There was a main effect of pulse-type ($F(1.27,13.967) = 34.559$, $p < 0.001$) and reaction time bin ($F(1.477, 16.244) = 1256.912$, $p < 0.001$), and a significant interaction ($F(1.943, 21.37) = 29.578$, $p < 0.001$). Planned contrasts indicated that post-pulse reaction times were significantly faster following a positive pulse ($F(1,11) = 25.88$, $p < 0.001$) and significantly slower following a reverse pulse ($F(1,11) = 15.843$, $p = 0.002$), compared to the gap pulse condition. By definition, the main effect of reaction time bin is accounted for by the increases in reaction times in later reaction time bins. The interaction was characterised with separate repeated measures ANOVAs for each pulse condition, using the mean post-pulse reaction time baselined by the gap pulse condition. Reaction time bin was the only factor in these analyses. There was a significant increase in the speeding effect of positive pulses as post-pulse reaction time increased ($F(1.399,15.384) = 25.798$, $p < 0.001$). There was also a significant increase in the slowing effect of reverse pulses ($F(3,33) = 5.967$, $p = 0.002$) as post-pulse reaction times increased.

4. Binned Analysis of Pulse Effect on Reaction Time Using Paired Sample t-tests

Table 6. Paired Sample t-tests of Reaction Time Pulse Effect Across Reaction Time Bins

Comparison	Mean Difference	SE	t	Sig. (2-tailed)
Positive - Bin 1	-0.001	0.004	-0.225	0.826
Positive - Bin 2	-0.044	0.012	-3.766	0.003
Positive - Bin 3	-0.090	0.016	-5.753	0.000
Positive - Bin 4	-0.094	0.018	-5.148	0.000
Reverse - Bin 1	0.006	0.004	1.560	0.147
Reverse - Bin 2	0.020	0.006	3.392	0.006
Reverse - Bin 3	0.027	0.007	4.016	0.002
Reverse - Bin 4	0.022	0.006	3.642	0.004

5. Mixed Effects Analysis of The Pulse Effect on the CPP and LRP

There was a main effect of pulse-type on the slope of the CPP measured in the window 200:500 ms after pulse onset ($F(2,25942.129) = 52.755$, $p < 0.001$). Pairwise comparisons showed that, compared to the gap pulse condition, the slope of the CPP was significantly greater following a positive pulse ($p < 0.001$). There was no significant difference between the reverse pulse and gap pulse conditions ($p = 0.823$).

There was a main effect of pulse-type on the slope of the LRP measured in the window 200:500 ms after pulse onset ($F(2,24609.272) = 17.705$, $p < 0.001$). Pairwise comparisons showed that, compared to the gap

pulse condition, the slope of the LRP was significantly greater following a positive pulse ($p < 0.001$). There was no significant difference between the reverse pulse and gap pulse conditions ($p = 0.175$).

† There was a significant difference between the reverse and no pulse conditions in the reported analyses. However, it is unsurprising that the comparison between the gap pulse and reverse pulse was not significant, since both reduced the slope of the LRP and CPP on correct response trials (see Figure 5.8.). The purpose of this analysis was to verify that these signals responded to the pulses, independent of the no pulse slicing technique. The significant effect of the positive pulse still illustrates this point.

6. **Mixed Effects Analyses of The CPP As a Single Accumulator**

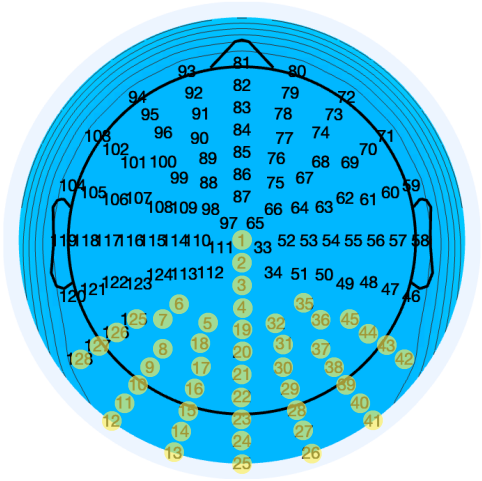
A mixed effects analysis using pulse-type as the only predictor was conducted on the CPP amplitude in the window 250:550 ms. The amplitude had been baselined in all conditions in the pulse-locked window -50:50 ms. On correct response trials, there was a main effect of pulse-type ($F(2,25951.429) = 94.371$, $p < 0.001$). Compared to the gap pulse condition, there was a significant increase in CPP amplitude following a positive pulse ($p < 0.001$). There was no significant difference between gap and reverse pulse trials ($p = 0.62$). There was no main effect of pulse-type on error response trials ($F(2,8172.273) = 1.807$, $p = 0.164$).

This analysis was repeated on the CPP slope measured from 150:350 ms. On correct response trials, there was a main effect of pulse-type ($F(3,25952.462) = 320.36$, $p < 0.001$). Compared to the gap pulse condition, there was a significant increase in CPP slope following a positive pulse ($p < 0.001$). There was no difference between gap and reverse pulse trials ($p = 0.754$). There was also a main effect of pulse-type on error response trials ($F(2,8176.816) = 6.984$, $p < 0.001$). Compared to the gap pulse condition, there was a significant increase in CPP slope following a positive pulse ($p < 0.001$) or reverse pulse ($p = 0.004$).

† Again, the non-significant comparisons between gap and reverse pulses is unsurprising and does not change the overall interpretation of these analyses. These analyses were conducted to justify the follow-up analyses of the absolute change in the CPP by demonstrating that there was a main effect of Pulse-Type.

Appendix 5.3 - Electrode Selection

SSVEP



The pool for selecting SSVEP electrodes is highlighted above. The individual electrode selections for each participant are listed below.

11, 10

23, 24

23, 22

22, 39

23, 22

23, 24

30, 31

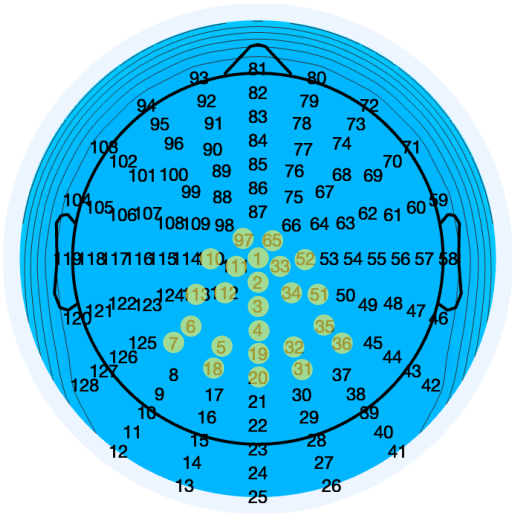
23, 22

24, 27

16, 23

10, 9

CPP



The pool for selecting CPP electrodes is highlighted above. The individual electrode selections for each participant are listed below.

4, 32, 19

2, 3, 4

19, 32, 4

5, 20, 19

2, 3, 4

3, 4, 19

2, 3, 4

2, 3, 4

20, 31, 19

3, 4, 19

3, 4, 19

20, 19, 4

LRP

There was no selection pool for the LRP. Participants electrodes (shown below) were selected from all 128 electrodes.

Left Hemisphere	Right Hemisphere
110, 114	53, 52
107, 108	63, 64
114, 110	52, 33
110, 109	53, 33
108, 107	67, 62
111, 97	55, 54
115, 108	52, 54
98, 109	67, 68
108, 107	64, 63
99, 88	63, 68
106, 107	64, 62

6.

The Influence of Expectation in Perceptual Decision-Making

No perceptual decision is made in the absence of any prior information (Bar, 2004; Gold & Stocker, 2017; Simoncelli, 2003; Summerfield & de Lange, 2014). For this reason, building an understanding of how expectations are encoded and integrated in perceptual decisions is a critical goal in cognitive science (Platt, 2002; Talluri et al., 2021). In perceptual decisions, sensory events elicit a cascade of activity distributed across the sensorimotor hierarchy; as sensory data percolates through processing streams, competing perceptual hypotheses are evaluated, and a response is prepared. There is a consensus amongst cognitive scientists that the brain exploits priors at some stage(s) in this decision process, but the question of how exactly this is achieved is a subject of ongoing debate (Friston, 2010; Heeger, 2017; Lochmann & Deneve, 2011; Pouget et al., 2013; Summerfield & de Lange, 2014; Teufel & Fletcher, 2020). Some have sought to address this question with a mechanistic account of the influence of priors in perception by analysing behavioural and neurophysiological responses to simple perceptual decisions within the framework of sequential sampling models (Gold & Stocker, 2017; Summerfield & Blangero, 2017).

The Drift Diffusion Model and The Influence of Expectation

When participants are provided with information about which of two decision alternatives are more likely, they exhibit biased responses favouring the more probable alternative with faster responses when the stimulus matches expectations, and slower responses to the unexpected stimulus (Summerfield & de Lange, 2014). The Drift Diffusion Model offers two candidate mechanisms to explain this effect (Bogacz et al., 2006; Diederich & Busemeyer, 2006; Ratcliff & McKoon, 2008). Expectations can bias the decision process by shifting the starting point closer to the bound associated with the more probable perceptual hypothesis (i.e. starting point bias; Figure 6.1.A). Alternatively, or in addition, the expectations can bias the rate of accumulation towards the expected choice, such that expected decisions are expedited and decisions in favour of the less probable alternative are protracted (i.e. drift bias; Figure 6.1.B).

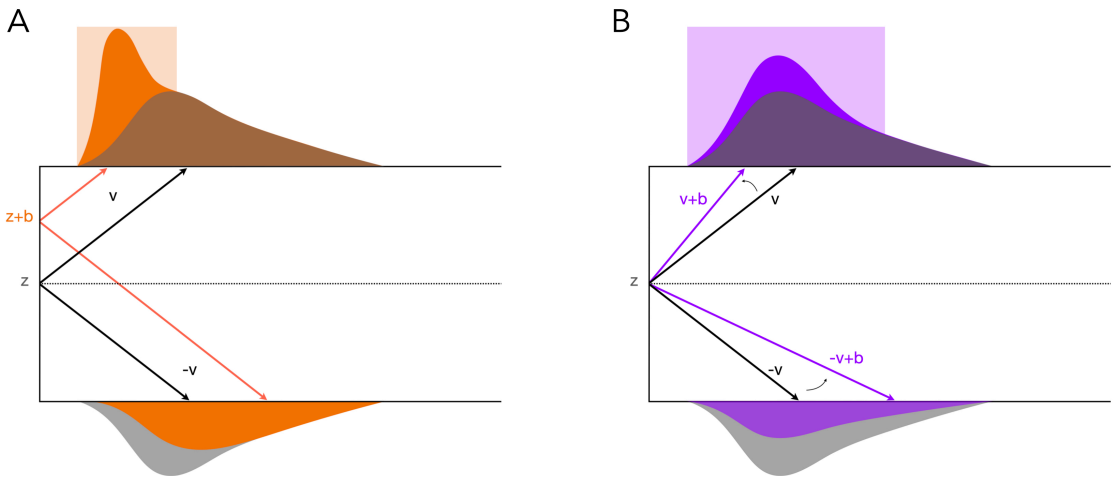


Figure 6.1. *Two accounts of the influence of priors on the decision process. A) The expectation is represented by an offset (b) in the starting point (z) of the decision variable, with no change in the drift rate (v). The shift closer to the bound mapped to the expected alternative is primarily evident in a significant increase in early correct responses and decrease in early error responses. B) The expectation is instantiated in a drift bias (b) which pushes the drift rate towards the expected alternative, with no change in starting point. This adjustment also speeds correct responses and slows error responses, but it has a more persistent effect on the reaction time distributions. The shaded regions of the correct response reaction time distributions highlights the portion of the trial where each adjustment is expected to exert an influence.*

In the model, a starting point bias is instantiated as a preparatory shift reflecting the conditional asymmetry, while a drift bias directly changes the trajectory of the decision variable, prejudicing the drift criterion that segregates the interpretation of sensory samples as evidence for one alternative or the other (see Figure 6.S1 in the Chapter Appendix; Mulder et al., 2014). In model simulations, this difference is most apparent in the leading edge of the reaction time distribution for the expected response. A starting point shift skews the distribution, with a heavy concentration of early responses; conversely, a small or moderate drift bias has little effect on the leading edge of the expected reaction time distribution, but a more prolonged effect overall (Forstmann et al., 2016; Ratcliff & McKoon, 2008). The accounts also differ in their response to varying evidence strength; a starting point shift is entirely evidence-independent, but its relative influence on the decision

variable will increase as the strength of the evidence decreases. Similarly, for a constant strength of evidence, the cumulative influence of a drift bias increases for longer periods of accumulation. However, the fixed effect of a starting point bias is expected to diminish as time elapses, as its relative contribution to the path of the decision variable is washed out by the increasing influence of accumulated evidence and noise. Although the characteristics of each adjustment described above may be readily apparent in model simulations, in empirical datasets these behavioural effects may be obscured by the interaction of multiple decision parameters, particularly if they have a more dominant influence on the decision dynamics. This is compounded by the emphasis on model parsimony, which rewards models that can adequately capture a prior modulation with one parameter, even if there may be more subtle dynamics at play.

Studies applying the Drift Diffusion Model to determine the mechanisms responsible for expectational effects have generally supported the starting point bias account (Leite & Ratcliff, 2011; Mulder et al., 2012; Ratcliff & McKoon, 2008), although this is not unanimous (Dunovan et al., 2014; Hanks et al., 2011; Kelly et al., 2021). To verify these efforts to model the implementation of prior-based decision-making, researchers have sought neurophysiological evidence of expectational effects at each key stage of the sensorimotor hierarchy: sensory encoding, evidence accumulation, and response preparation.

Expectation and Sensory Processing

The Drift Diffusion Model does not define the sources of drift rate bias or starting point shifts, so it is agnostic as to whether the adjustments occur at the sensory, decision, or motor level, or some combination of these. For example, drift biases could conceivably originate at the motor level, taking the form of an evidence-independent imbalance in the rate of preparation for each alternative action without impacting on upstream processing levels. Equally, drift biases could originate at a sensory level, such that the representation of the evidence itself is biased towards a particular perceptual hypothesis, with knock-on effects for the rate of evidence accumulation at the decision level and the rate of preparation at the motor level. There is mounting neurophysiological evidence of expectation effects in sensory processing (de Lange et al., 2018; Teufel & Fletcher, 2020), but there is some debate as to whether these sensory modulations are epiphenomena, echoing the conclusion of the decision process, or if they play an important role in the formation of the decision.

There is evidence that predictive cues and stimulus associations can activate neural populations tuned to the stimulus feature being predicted, ranging from direction-selective MT activity for predicted motion (Albright, 2012; Schlack & Albright, 2007; Shulman et al., 1999) to the activation of the Fusiform Face Area or Parahippocampal Place Area for expected face and house stimuli, respectively (Egner et al., 2010; Puri et al., 2009; Trapp et al., 2016). In a series of fMRI and MEG decoding studies Kok et al. (2012, 2013, 2014, 2017) provided evidence that this activity encodes representations of the expected stimulus features before stimulus onset and alters the representation of the subsequently presented physical stimulus. For example, Kok et al. (2012) found that expectations dampened responses to the stimulus in primary visual cortex, but the

classification accuracy of the decoded orientation was significantly improved compared to unexpected gratings. This observation provides a potential solution to an apparent contradiction in the expectation-driven sensory modulation account. Many studies have reported that the aggregate responses to expected stimuli in associated sensory areas are suppressed in comparison to unexpected or neutral stimuli (i.e. expectation suppression; Feuerriegel et al., 2021a; Summerfield & de Lange, 2014; Walsh et al., 2020), but it is counterintuitive that dampened sensory responses would lead to increased evidence accumulation for the expected stimulus. However, Kok et al. (2012, 2014) reported that expectation suppression disproportionately targeted voxels tuned away from the expected stimulus features, which suggests that selective inhibition may actually serve to sharpen the representation of the expected stimulus by quelling, or pre-emptively suppressing, incongruent responses that feed alternative perceptual hypotheses or inject noise into the signal. This increase in the signal-to-noise ratio would be expected to increase the drift rate, favouring the expected sensory evidence. Alternatively, Feuerriegel et al. (2021b) propose that these pre-stimulus sensory templates are an early source of sensory evidence for pre-evidence accumulation, leading to a starting point bias. The authors predict that this should manifest in the early onset of the CPP and an increased amplitude at evidence onset compared to neutral expectation conditions.

Expectation and The Decision Process

Converging lines of neurophysiological research in monkeys have provided strong evidence that the decision state is readily detectable in a distributed network of regions associated with the preparation of a motor response (discussed in Chapter 4). Neural correlates of the starting point bias have been identified in changes of baseline firing rates in a number of oculomotor areas in saccadic response tasks, including the superior colliculus (Basso & Wurtz, 1997, 1998; Dorris & Munoz, 1998) and area LIP (Platt & Glimcher, 1999; Rao et al., 2012). The potential importance of this elevated baseline activity was illustrated by Hanks et al. (2006), who used microsimulation to slightly enhance activity in LIP neurons while monkeys performed random dot motion discrimination. Artificially raising the baseline firing rates of LIP neurons produced exactly the same behavioural changes as prior information induces: more frequent and faster responses in the associated direction and slower responses for the alternative response.

Researchers have captured equivalent responses in humans based on subsequent evidence that the decision state is also encoded in human motor preparation signals, including the lateralised readiness potential and beta-band oscillatory activity (Kelly & O'Connell, 2013; O'Connell et al., 2012; Siegel et al., 2011; Wilming et al., 2020). Beta-band activity over premotor cortex has been found to exhibit the same evidence-dependent ramping observed in other neural correlates of the decision variable, which predicts choice seconds before the response is made and culminates at a stereotyped threshold immediately before the response is executed (Donner et al., 2009; Kubanek et al., 2013). O'Connell et al. (2012) found that, like the CPP (discussed in Chapter 4), the slope of contralateral beta activity increased across evidence presentation, indicating that both

signals encoded a representation of the integrated sensory evidence. De Lange et al. (2013) found that lateralised beta-band activity exhibited a baseline shift in response to a predictive cue presented in advance of evidence onset. This finding nicely complemented a previous observation that the level of this preparatory beta desynchronisation scales with the level of uncertainty about the target of an upcoming reaching movement (Tzagarakis et al., 2010). Given the evidence that the ramping activity in motor preparation signals reflects the accumulated evidence in the decision process (Donner et al., 2009; O'Connell et al., 2012; Gold & Shadlen, 2007), the changes in pre-evidence baseline activity observed in response to predictive cues correspond to a starting point shift in the decision variable (see Figure 6.S2 in the Chapter Appendix).

Much of this work examining expectation effects on decision signals has supported the starting point bias account. However two key studies have provided reason to not dismiss the idea that priors may exert a more persistent influence on the accumulation of evidence (Hanks et al., 2011; Kelly et al., 2021). On average, strong evidence leads to fast and accurate responses (Ratcliff & Smith, 2004), so if the decision-maker finds themselves still deliberating after a relatively long period of time, it is likely because the evidence is too weak to allow for a commitment to either decision alternative. This also means that a decision made after a long period of time is less likely to be correct. Therefore, when a participant does not know how strong the evidence being presented is, the period of time they spend deliberating on their response is a good proxy for the reliability of the evidence. In a study of non-human primates, Hanks et al. (2011) provided evidence that the brain can exploit this property of the decision process to dynamically adjust the relative influence of prior probability and sensory evidence in proportion to its estimated reliability. On trials with fast responses, the sensory evidence is likely reliable, so expectations can be neglected. Conversely, on trials with slow responses, the sensory evidence is likely uninformative, so priors can be increasingly leaned upon to guide the decision process (Figure 6.2.A).

Hanks et al. suggested that a time-dependent influence of priors could be implemented with a 'dynamic bias signal' which adds or subtracts an increasing quantity to the decision variable based on the relationship between accuracy and elapsed decision time. Model fits of human and monkey performance on a random dot motion discrimination task, with block-by-block biases in the frequency of left/right targets, showed that the dynamic bias signal was better able to account for the behavioural data than a starting point bias. Consistent with previous studies, Hanks and colleagues also found that the firing rates of LIP neurons were elevated before stimulus onset when the prior favoured their preferred response. Critically, higher prior probability also led to an acceleration in the build-up rate of LIP activity over the course of the trial and this effect was stronger for weaker motion coherence, exactly as the dynamic bias account predicted. Finally, the authors showed that the combination of a model fitted to neutral prior trials and a dynamic bias signal estimated from the neural data could account for behaviour on trials with imbalanced prior probabilities (Hanks et al., 2011).

Hanks et al.'s estimate of the dynamic bias is not zero at evidence onset, consistent with their observation of prior-induced changes in LIP baseline activity before the evidence onset. So the dynamic bias signal effectively represents a single mechanism for both a starting point offset and a drift bias (Martinez-Rodriguez et al., 2020).

This idea of a shared mechanism is supported by the observation that variation in pre-stimulus LIP activity was correlated with the firing rate during evidence presentation on a trial-to-trial basis. A dynamic drift bias is an interesting proposal because it predicts the exact opposite of the starting point bias account, where the influence of prior expectation decreases as more evidence is accumulated. However, there is little empirical work testing the hypothesis that priors become increasingly influential at later periods of the trial. There is evidence that pulses of evidence with later onsets are less influential, both behaviourally and in pulse-evoked LIP responses (Huk & Shadlen, 2005; Kiani et al., 2008), but these experiments did not manipulate sensory expectations. The study presented in this chapter looked to test this hypothesis directly.

In another key study, Kelly et al. (2021) developed an account of the influence of priors in perceptual decision-making by constructing a model where parameters were directly constrained by empirically-observed, neurophysiological indices, and then testing this model against independent neural data. Specifically, they investigated if the resulting neurally-informed (NI) model could capture subtleties in the prior-induced adjustments that may go undetected by the standard Drift Diffusion Model. Their NI model differed by incorporating a time-dependent urgency signal which is added at the motor level. This inclusion was supported by the observation that beta activity commenced its build-up toward threshold well before evidence onset, a trend that was accompanied by a reduction in CPP amplitude for longer reaction times (also reported by Steinemann et al., 2018), consistent with a collapsing bound effect. The NI model's bound, starting point, and motor-execution delay were tethered to mu-beta amplitude at response, pre-evidence mu-beta lateralisation, and the onset of motor execution potentials, respectively. Importantly, the quantitative constraints on these parameter values offset the loss of parsimony, ensuring that these additions did not result in unjustifiable flexibility, and left the model better positioned to test for more subtle effects like drift bias.

Kelly et al. examined the influence of priors in three random dot motion discrimination conditions: two conditions had long response deadlines and differed in having high ('Easy') or low ('LoCoh') motion coherence; a third condition had a short deadline and high coherence ('Deadline'). In all conditions, mu-beta lateralisation was observed before evidence onset, reflecting preparation of the cued response, which the authors point out is a commonly observed signature of a dynamic urgency component. This motor preparation was reflected in the fits of both the NI model and the Drift Diffusion Model as a starting point bias. There was also an enduring effect of the cue on accuracy across all reaction time bins. This is more readily explained by a drift bias than a starting point bias, which is expected to primarily influence the leading edge of the reaction time distribution (see Figure 6.1.A). While both models produced positive drift bias estimates, model comparison scores indicated that the inclusion of this additional parameter was only statistically justifiable in the NI model. The inclusion of a drift bias was further supported by an effect of the cue on the response-locked CPP amplitude in the Deadline condition. Specifically, the CPP amplitude at response was inversely proportional to the prior probability of the cued response in the Deadline condition, but there was no difference in CPP amplitude across cue conditions in the LoCoh condition. Kelly et al. attributed this effect to the interplay of urgency and drift bias. This complex pattern, shown in Figure 6.2.B, is a good illustration of why drift bias can be difficult to detect. Overall, the NI model with the drift bias provided a better fit to the behavioural data than the Drift

Diffusion Model. However, it is unclear from what level of the decision process this drift bias arose, and Kelly et al. did not include a neurophysiological marker of sensory encoding. The study presented in this chapter sought to replicate some of these novel results and address the question of a sensory modulation using the steady-state visual evoked potential (SSVEP).

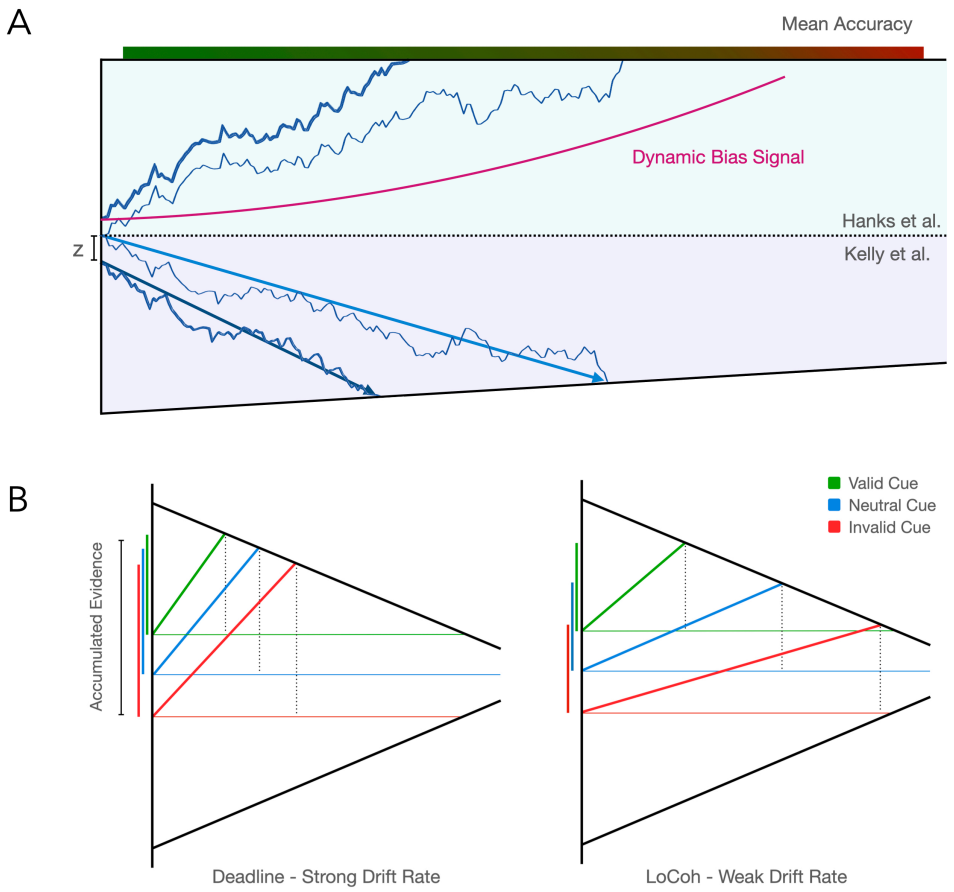


Figure 6.2. *A) A comparison of the effects of priors according to Hanks et al. (2011) and Kelly et al. (2021). According to Hanks et al. (top), a dynamic bias signal is added to the accumulated evidence (thin line), which increasingly pushes the decision variable (thick line) towards the bound associated with the expected alternative as time elapses. The strength of the dynamic bias signal scales with the average probability of a correct response across the duration of the trial. The dynamic bias signal was estimated to be non-zero at evidence onset, effectively implementing a starting point bias. According to Kelly et al., priors elicit a starting point bias (z), and a drift rate bias (dark blue). Their model also incorporates urgency, which symmetrically collapses the bounds across the trial. In both cases, decision formation is increasingly pushed towards the expected alternative as the decision unfolds. B) Kelly et al. found that the response-locked amplitude of the CPP was smaller for valid cues and larger for invalid cues in the Deadline condition, but not in the LoCoh condition. This suggests that the quantity of evidence needed to terminate the decision process (coloured bars on left of each illustration) varied across cues in the Deadline condition, but was constant in the LoCoh condition. In the Deadline condition, the starting point bias had a significant impact on the amount of evidence needed to reach the bound, but the collapsing bound and drift bias had little impact because the drift rate was strong. Conversely, the weak drift rate in the LoCoh condition amplified the effect of the drift bias, but also allowed the collapsing bound to shape the quantity of evidence needed. These mechanisms had opposite effects, resulting in no difference in the accumulated evidence across cue conditions.*

The studies presented by Hanks et al. (2011) and Kelly et al. (2021) are outliers. In contrast to the overwhelming majority of the evidence reviewed so far, they endorse a drift bias. However, they deserve more weight than their arithmetic contribution to the debate because, unlike many of the previous studies, these accounts ground model parameters in both behavioural and neural data. The Drift Diffusion Model allows prior expectations to be instantiated in a starting point bias, but there is no agreed-upon formula for calculating the magnitude of that offset a priori, instead the drift rate and starting point parameters are typically free to vary for each model fit. It is interesting that when Hanks et al. and Kelly et al. constrained model parameters based on neurophysiological indices of the decision process, they arrived at very similar accounts.

Summary

The present study was designed to simultaneously capture the evolution of the decision process across three key stages of the sensorimotor hierarchy by tracing the dynamics of well-characterised electrophysiological signatures of sensory encoding (the steady-state visual evoked potential; SSVEP), evidence accumulation (the centroparietal positivity; CPP), and motor preparation (mu-beta desynchronisation; MB). Specifically, the aim was to identify characteristic behavioural markers of biases in starting point and drift rate and to detect plausible neurophysiological manifestations of these adjustments at each of these decision stages. Three central questions were investigated: 1) Do expectations influence the encoding of sensory evidence either before evidence onset or during evidence presentation?; 2) Is there consistent behavioural and neurophysiological evidence of a starting point bias or drift bias or both?; 3) Does the influence of the expectation on sensory encoding or evidence accumulation change across the duration of the trial? These questions were addressed with a cued contrast discrimination task, incorporating evidence pulses with unpredictable onset times. It has already been demonstrated in Chapter 5 that participants accumulate evidence on this task. It was hoped that the interaction of prior-informed evidence integration and pulses could shed light on questions 2 and 3. If the cue provokes a drift bias, the response to a pulse of evidence should be enhanced when it is congruent with the cue compared to when there is no cue or the pulse conflicts with expectations. Similarly, if the drift bias grows over the course of the trial, an early pulse should elicit a smaller response than a later pulse.

Methodology

The analyses reported in this chapter use data collected in the same experiment described in Chapter 5. To avoid repetition, only aspects of the data analysis that were either unique to this chapter or central to the interpretation of the results are described in this section.

Participants

12 adults participated in this study (5 female, age range: 18-39, M = 25.2 ± 6.7). However, a reliable SSVEP could not be established for one participant so they were excluded from the SSVEP analyses (n = 11), but retained for all other analyses (n = 12). The SSVEP was considered reliable if it exhibited a clear divergence in the signals representing the high-contrast and low-contrast gratings at evidence onset. All participants reported normal or corrected-to-normal vision, no history of migraine or bad headaches, no history of epilepsy, and no sensitivity to flashing light. Participants provided informed written consent prior to testing and were paid a gratuity of €10 per hour of participation as compensation for their time. Participants received an additional €20 for completing all testing sessions. All procedures were approved by the Trinity College Dublin School of Psychology Ethics Committee and were in accordance with the Declaration of Helsinki.

Data Analysis

As discussed in the methodology section of Chapter 5, the no-pulse data were divided into 'slices' in which reaction time restrictions were imposed to match those applied to each of the pulse onset conditions. However, since the data for each pseudo onset time were selected from the full no-pulse dataset, this meant that a given trial could be included in several or all slices. The inclusion of these replicated trials may raise concerns about falsely inflating the power of those statistical analyses using a mixed effects approach. To address any such concerns, all analyses that used the sliced no-pulse data were repeated, removing all no-pulse trials and using the gap pulse as the control condition (see Chapter Appendix). This did not meaningfully change any of the results. This sliced no-pulse dataset was used as the control condition for all analyses of the pulse effect reported in this chapter. Any analysis that did not include pulse as a factor, did not use this sliced dataset. All analyses of cue-locked data included all trials, independent of the validity of the response, but evidence-locked, pulse-locked, and response-locked analyses included only correct and error responses. The average valid response rate across participants was 97.5%, so it was not expected that these inclusions/exclusions were an important factor in any of the results. In the cue-locked epoch, there is no meaningful distinction between valid cues and invalid cues, they only become valid/invalid after evidence onset. For this reason, cue-locked

analyses pooled trials from valid and invalid cue conditions into a 'cued' condition. On a cued trial, there was a cued stimulus (consistent with the cue direction) and an un-cued stimulus (the opposite of the cued direction). For comparison with the cued condition, the stimuli in neutral cue trials were labelled according to what would eventually be the correct target.

Most of the statistical analyses were based on mixed effects modelling. This was particularly important in this study because there was a small sample, but a large dataset for each participant. Mixed effects analyses exploit these trial numbers to reduce the risk of Type II error. As mentioned in the methodology section of Chapter 5, there is no consensus on how to test for fixed effects mixed effects models. However, it has been recommended that studies with small sample sizes estimate the degrees of freedom using a Satterthwaite approximation and compute an F-statistic (Kuznetsova et al., 2017; Luke, 2017). This approach was adopted for all mixed effects analyses. A random intercept was included in all models to account for the repeated-measures nature of the data. For some analyses a mixed effects approach was inappropriate and a repeated-measures ANOVA was used instead. In these cases, where Mauchly's test indicated that the assumption of sphericity had been violated, a Greenhouse-Geisser correction was applied to the degrees of freedom. The significance of all factors included in each analysis is reported. Unless stated otherwise, all main effects were investigated with uncorrected pairwise comparisons.

EEG Analysis

All of the EEG preprocessing steps were identical to Chapter 5. The primary difference in this chapter is that mu-beta oscillatory activity (8-30 Hz) was used to measure motor preparation instead of the lateralised readiness potential (LRP) used in Chapter 5. As an ERP, the LRP offers excellent temporal resolution, which was required for the timing analysis reported in the previous chapter. In contrast, mu-beta activity, which also represents motor preparation in humans (Donner et al., 2009; de Lange et al., 2013), is extracted using a Fourier transformation. Because this process estimates oscillatory activity using a window of EEG data, the temporal precision of the signal is degraded, but the signal is also less noisy than an ERP. Most importantly, however, mu-beta provides a distinct read-out of motor preparation for each of the decision alternatives, while the LRP only offers information about relative preparation. This was deemed critical for capturing the cue effect, which may differently affect motor preparation for the cued and uncued response, and to detect any influence that affected preparation for each response alternative equally (e.g. dynamic urgency signals).

The procedure for extracting oscillatory signals with the Short Time Fourier Transform (STFT) was reported for the SSVEP in Chapter 5. Here, the exact same STFT procedure was used to extract both the SSVEP signals associated with the tilted gratings (20 Hz and 25 Hz) and mu-beta activity. The CPP and SSVEP electrode selections were retained from Chapter 5. Left and right hemisphere mu-beta electrodes were separately selected for each participant from a candidate pool (see Chapter Appendix). Electrodes were ranked based on two criteria: 1) the difference in amplitude when a candidate electrode was contralateral and ipsilateral to the

response; 2) the slope of activity at the candidate electrode in the 400 ms preceding the contralateral response. The response-locked topography of mu-beta activity was consulted to confirm the top-ranked electrodes were consistent with the sources of differential activity preceding a response. 2-4 of the top ranked electrodes were chosen for each hemisphere for each participant.

Results

Effects of Cues and Pulses on Reaction Time

The first analysis addressed the influence of the cues on reaction times in correct and error response trials (Figure 6.3). Reaction time data were pooled across pulse-type and pulse onset conditions and assessed with a repeated-measures ANOVA. There was no main effect of Cue Condition ($F(2,22) = 1.679$, $p = 0.21$) or Accuracy ($F(1,11) = 0.039$, $p = 0.847$), but there was a significant interaction ($F(1.125,12.370) = 37.399$, $p < 0.001$). Separate analyses of correct and error reaction times were conducted to explain this interaction. There was a main effect of cue condition on correct response trials ($F(2,22) = 27.226$, $p < 0.001$) with reaction times significantly reduced in the valid cue condition ($\Delta = -41 \pm 39$ ms, $p = 0.039$) and significantly increased in the invalid cue condition ($\Delta = 94 \pm 39$ ms, $p < 0.001$) compared to the neutral cue condition. There was also a main effect of cue condition on error response trials ($F(2,22) = 16.003$, $p < 0.001$) with reaction times significantly slower following a valid cue ($\Delta = 86 \pm 52$ ms, $p = 0.004$), but the reduction in reaction time on invalid cue trials was not significantly different from the neutral cue condition ($\Delta = -45 \pm 49$ ms, $p = 0.067$).

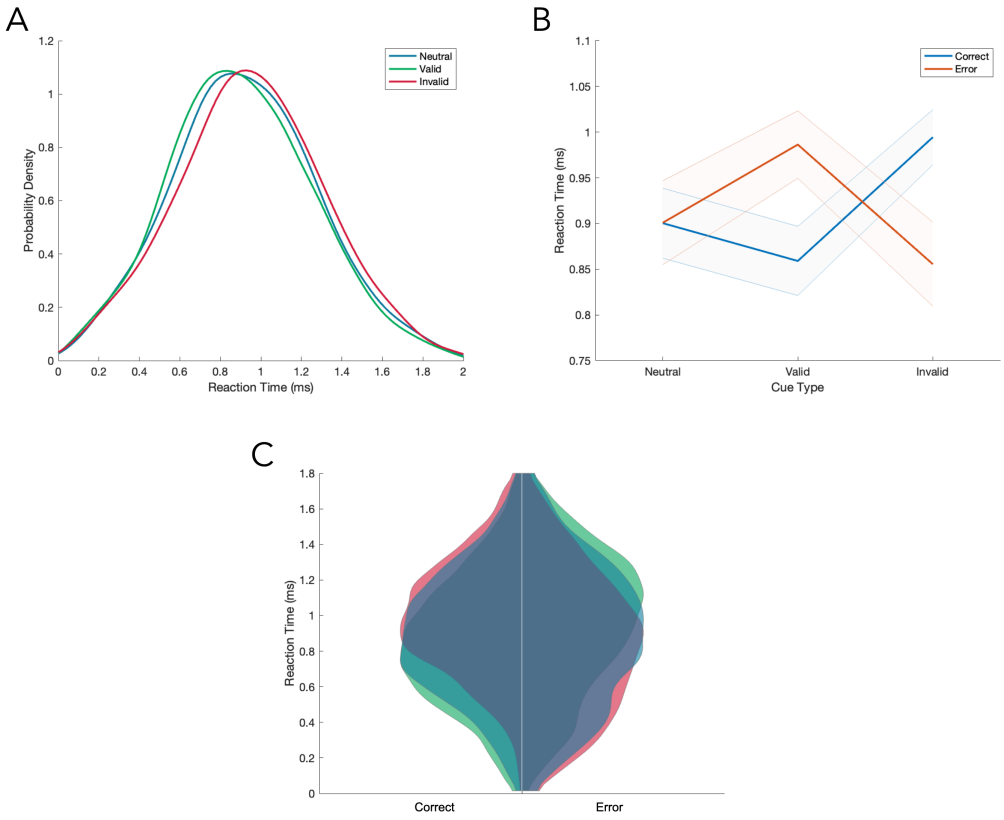


Figure 6.3. *The cue effect of reaction time. A) The grand average reaction time distribution for each cue condition. B) There was a significant interaction in the effect of the cue across correct and error trials. The shaded region represents the standard error of the mean. C) Violin plots showing the distribution of reaction times for each cue condition. Correct responses are shown on the left of the plot and error responses are shown on the right. The cue conditions retain the same colour codes from (A).*

Examining the effects of pulses on reaction times, there was a main effect of Accuracy ($F(1,11) = 16.815$, $p = 0.002$) and Pulse-Type ($F(2.025,22.28) = 9.736$, $p < 0.001$), and a significant Accuracy by Pulse-Type interaction ($F(3,33) = 8.222$, $p < 0.001$). The main effect of Accuracy reflected the fact that reaction times were faster on correct response trials ($\Delta = -46 \pm 24$ ms, $p = 0.002$). Follow-up comparisons for the main effect of Pulse-Type indicated that positive pulses led to significantly faster responses than no-pulse trials ($\Delta = -33 \pm 23$ ms, $p = 0.01$); there was no significant difference between reaction times on gap ($\Delta = 4 \pm 18$ ms, $p = 0.611$) or reverse pulse trials ($\Delta = 4 \pm 19$ ms, $p = 0.628$) and those in the no-pulse condition. Correct and error responses were analysed separately to characterise the interaction. There was a main effect of Pulse-Type for correct responses ($F(3,33) = 26.631$, $p < 0.001$) due to significantly faster responses following positive pulses ($\Delta = -50 \pm 23$ ms, $p < 0.001$) and significantly slower responses following reverse pulses ($\Delta = 29 \pm 16$ ms, $p = 0.003$), compared to the no-pulse condition. There was no significant difference between gap pulse and no-pulse reaction times ($\Delta = 13 \pm 18$ ms, $p = 0.13$) and no main effect of pulse on reaction times for error responses ($F(1.964,21.604) = 0.923$, $p = 0.411$).

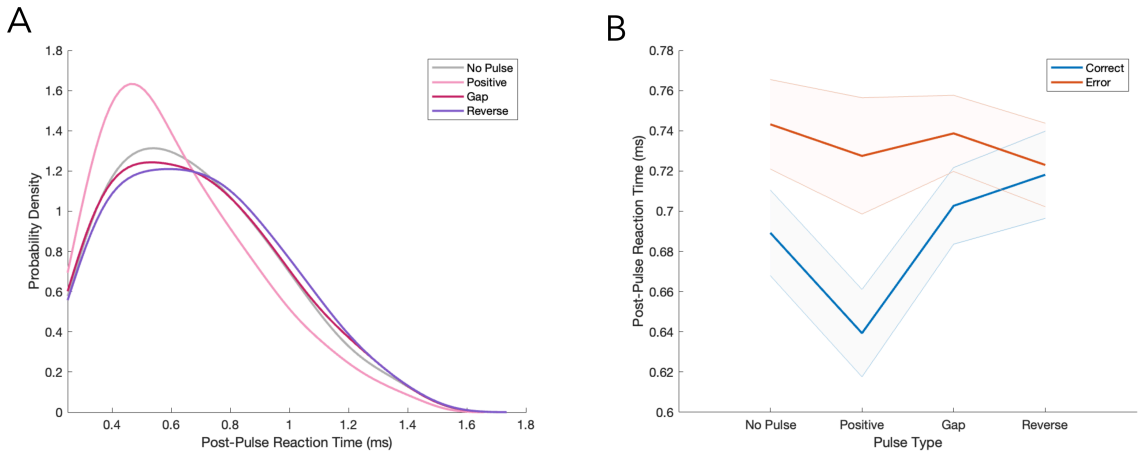


Figure 6.4. *The pulse effect on reaction time. A) The grand average post-pulse reaction time distribution for each pulse condition. The response had to be at least 250 ms after the pulse to be included. The grey line shows the sliced no-pulse condition. B) Mean post-pulse reaction times across pulse-types for correct and error response trials. The shaded region represents the standard error of the mean.*

In summary, valid cues led to faster correct reaction times and slower errors, while invalid cues led to slower correct reaction times, but not significantly faster errors. On all trials, reaction times were faster when a positive pulse occurred. Reaction times were slower on reverse pulse trials, but only when the response was correct. These results confirm the basic expectation that the cues and pulses influenced participant behaviour on this task. The observation that reverse pulses slowed down correct responses, shows that the pulses influenced the trajectory of the decision, even when that influence did not determine the response. It can also be noted that there does not appear to be a disproportionate number of fast responses in the valid cue condition, as predicted by the starting point bias account; the valid cue reaction time distribution only diverges from the neutral cue condition at approximately 400 ms (Figure 6.3.A). In addition, the effects of the cues are seen throughout the distributions of correct and error response reaction times (Figure 6.3.C).

A central hypothesis in this study was that the prior would shape the influence of the sensory evidence on the decision process, so the next analysis was conducted to determine if the cue moderated the influence of a pulse on reaction time (Figure 6.5). Since the previous analyses have demonstrated that the impact of a cue or a pulse varies depending on whether the response was correct or an error, these were analysed with separate mixed effects analyses. For correct response trials, there was a main effect of Pulse-Type ($F(3,47776.101) = 107.416$, $p < 0.001$) and Cue Condition ($F(2,47777.35) = 187.522$, $p < 0.001$), but there was no interaction ($F(6,47776.096) = 1.304$, $p = 0.251$). Compared to neutral cues, valid cues significantly reduced reaction times ($\Delta = -27 \pm 7$ ms, $p < 0.001$) and invalid cues significantly increased reaction times ($\Delta = 46 \pm 9$ ms, $p < 0.001$). In addition, positive pulses reduced reaction times ($\Delta = -49 \pm 9$ ms, $p < 0.001$), while gap ($\Delta = 10 \pm 9$ ms, $p = 0.031$) and reverse pulses ($\Delta = 29 \pm 9$ ms, $p < 0.001$) slowed reaction times. For error responses trials, there was a main effect of Cue Condition ($F(2,14198.446) = 95.99$, $p < 0.001$), but there was no main effect of Pulse-Type ($F(3,14192.946) = 1.544$, $p = 0.201$), and no interaction ($F(6,14192.347) = 1.778$, $p = 0.099$). Valid cues significantly increased reaction times ($\Delta = 48 \pm 13$ ms, $p < 0.001$) and invalid cues significantly reduced reaction times ($\Delta = -29 \pm 13$ ms, $p < 0.001$). The lack of a significant interaction in these analyses does not add support to the drift bias hypothesis as congruent cue-pulse pairings were expected to accelerate evidence accumulation and result in faster reaction times.

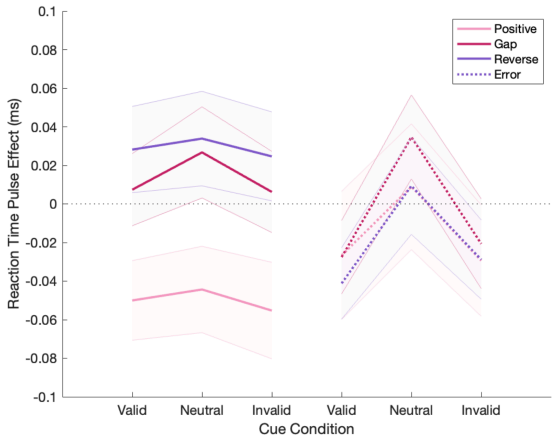


Figure 6.5. *The combined effects of cues and pulses on reaction time for correct (right) and error (left) trials. The pulse effect was defined as the difference in reaction time between the sliced No Pulse condition and each of the pulse conditions. The shaded region represents the standard error of the mean.*

It was also hypothesised that the influence of the prior on behaviour would change across the duration of the trial based on the suggestion that there is a dynamic shift in the relative influence of the sensory evidence and the prior as the trial progresses, such that the sensory evidence is less influential at later stages. If there was such a dynamic drift bias, the effect of a pulse of evidence may change depending on whether the pulse occurred earlier or later in the trial. However, at later stages in the trial, the decision variable is likely to be closer to the correct decision bound, since the participant has been exposed to positive evidence in the lead up to pulse onset and mean accuracy was ~70%. Indeed the delay between pulse onset and response ('post-pulse

reaction time') became smaller as pulse onset increased (see Figure 6.6.A). As illustrated in Figure 6.6.B, as the decision variable approaches the bound, the scope for pulses to impact on reaction time diminishes. Since the proximity of the decision variable to the bound can potentially mask the weighting of a pulse of evidence, a trend of positive pulses becoming less influential as pulse onset time increases is a feature of this paradigm and cannot itself offer any insight into the question of a dynamic drift bias. In order to control for the relationship between time and proximity to bound a baseline comparison condition was generated by 'slicing' the no-pulse data into five pseudo-onset times, which mirror the decline in post-pulse reaction times seen in the other pulse conditions (Figure 6.6.A). The sliced no-pulse condition was generated by iteratively imposing the reaction time restrictions applied to each pulse onset (reaction times must be 250 ms after the pulse onset time) to the same no pulse dataset (see Chapter 5 Methods section for further details).

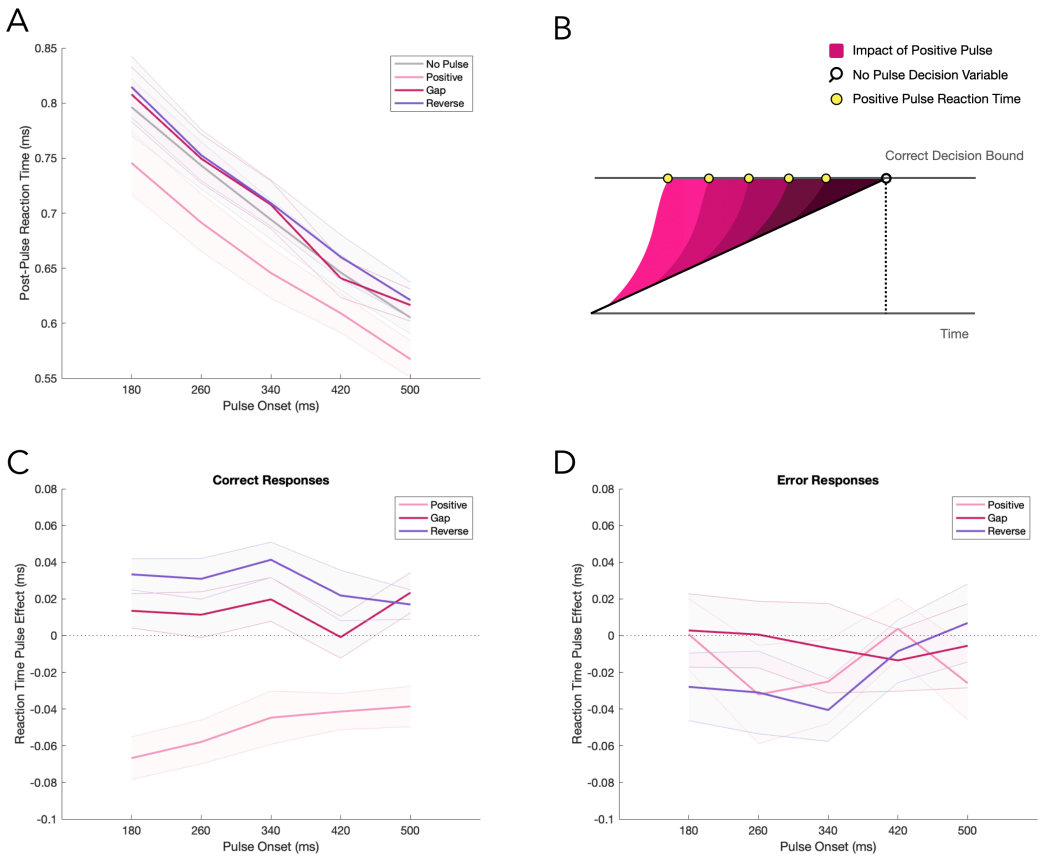


Figure 6.6. *The effect of the pulse on reaction time at different onset times. A) The time between pulse onset and the response decreases as the pulse onsets later in the trial, in all conditions. B) This can be explained by the average proximity of the decision variable to the bound at later onset times. An early positive pulse has the potential to substantially accelerate the decision variable towards the correct decision bound. However, as the onset time gets later, the average decision variable has gotten closer to that bound because it has been exposed to normal evidence in the interim. The mean pulse effect at each of the five onset times is shown for correct (C) and error (D) responses.*

For correct responses, there was a main effect of Pulse-Type ($F(3,47780.043) = 94.671$, $p < 0.001$), Pulse Onset ($F(1,47780.767) = 1461.987$, $p < 0.001$), and a significant interaction ($F(3,47780.033) = 4.664$, $p = 0.003$). Again, positive pulses reduced reaction times (Δ = -46±7 ms, $p < 0.001$), while gap (Δ = 9±7 ms, $p < 0.001$) and

reverse pulses (Δ = 29±7 ms, $p < 0.001$) increased reaction times. The main effect of Pulse Onset was characterised by running the analysis again with the interaction term excluded. Reaction times significantly increased as Pulse Onset increased ($F(1,47783.772) = 1482.37$, $\beta = 0.417$, $p < 0.001$). This was a necessary consequence of the trial inclusion criteria for pulse analyses (see Chapter 5 Methods). The interaction was interpreted using the model coefficients, which indicate if the relationship between a pulse condition and onset time significantly deviates from the relationship of the reference no-pulse condition. The decrease in reaction times associated with positive pulses reduced across pulse onsets ($\beta = 0.075$, $p = 0.012$), but there was no significant difference between the gap ($\beta = -0.021$, $p = 0.502$) or reverse pulse ($\beta = -0.019$, $p = 0.536$) conditions and the no-pulse condition.

For error responses, there was a main effect of Pulse-Type ($F(3,14196.429) = 3.241$, $p = 0.021$) and Pulse Onset ($F(3,14196.375) = 341.729$, $p < 0.001$), but no interaction ($F(3,14196.157) = 1.605$, $p = 0.186$). Reverse pulses significantly reduced reaction times (Δ = -15±12 ms, $p = 0.024$), but there was no significant difference between Positive (Δ = -7±14 ms, $p = 0.368$) or gap pulse (Δ = -2±13 ms, $p = 0.808$) conditions and the no-pulse condition. As mentioned in the previous analysis, reaction times necessarily increased as pulse onset increased ($F(1,14199.394) = 361.809$, $p < 0.001$).

The analysis was repeated with an additional measure taken to limit the potential influence of the proximity to the bound issue. The pulse effect on reaction time was expressed as a percentage of the sliced no-pulse post-onset reaction time. In other words, the pulse effect was measured as a proportion of the scope for a change in reaction time[1] (shades of pink in Figure 6.6.B). For correct responses, there was a main effect of Pulse-Type ($F(3,47780.639) = 80.076$, $p < 0.001$), but no main effect of Pulse Onset ($F(1,47789.228) = 1.676$, $p = 0.195$) and no interaction ($F(3,47780.481) = 1.199$, $p = 0.309$). For error responses, there was a no main effect of Pulse Type ($F(3,14200.018) = 2.503$, $p = 0.057$), no main effect of Pulse Onset ($F(1,14199.974) = 1.537$, $p = 0.215$), and no interaction ($F(3,14197.814) = 1.624$, $p = 0.182$).

In summary, it is difficult to definitively test for changes in the magnitude of the pulse effect on reaction times when the pulse is delivered at different stages of the trial because on average the decision process will be in different states at these different pulse onset times (Huk & Shadlen, 2005; Kiani et al., 2008). Here, estimating the impact of the pulse using subsets of the no-pulse trials and expressing the pulse effect relative to the no-pulse condition indicated that there was no significant change in the pulse effects across pulse onset times.

[1] For example, if there was a mean reaction time of 1000 ms in the No Pulse condition at the pseudo onset time of 500 ms, and the mean reaction time on positive pulse trials at the same onset was 850 ms, the positive pulse effect at 500 ms resulted in a 15% faster reaction time.

$$PE = \frac{(850 - 500) - (1000 - 500)}{(1000 - 500)} = -15\,\%$$

Effects of Cues and Pulses on Accuracy

Next, the influence of the cues and pulses on choice accuracy was assessed. A repeated measures ANOVA was conducted on the accuracy data to assess the influence of the cues and pulses (Figure 6.7). There was a main effect of Cue Condition ($F(1.165,12.82) = 51.93$, $p < 0.001$), and Pulse-Type ($F(1.851,20.361) = 75.737$, $p < 0.001$), and a significant interaction ($F(33.081,33.893) = 7.325$, $p < 0.001$). Accuracy was higher on valid cue trials ($\Delta = 15.6 \pm 5.7\%$, $p < 0.001$) and lower on invalid cue trials ($\Delta = -17.1 \pm 5\%$, $p < 0.001$), compared to the neutral cue condition. Positive pulses increased accuracy ($\Delta = 8.4 \pm 2.7\%$, $p < 0.001$), while gap ($\Delta = -3.1 \pm 1.9\%$, $p < 0.001$) and reverse pulses ($\Delta = -6.6 \pm 1.1\%$, $p < 0.001$) reduced accuracy, compared to the no-pulse condition. The interaction was investigated with separate ANOVAs for each of the pulses baselined by the no-pulse condition. There was a main effect of Cue Condition in the baselined positive pulse data ($F(2,22) = 10.062$, $p < 0.001$). The effect of the positive pulse was significantly greater in the invalid cue condition than the neutral cue condition ($\Delta = 6.3 \pm 5.6\%$, $p = 0.03$). The difference between valid and neutral cue conditions failed to reach significance ($\Delta = -4.1 \pm 4.2\%$, $p = 0.056$). There was no main effect of Cue Condition in the baselined gap pulse ($F(2,22) = 0.539$, $p = 0.591$) or reverse pulse data ($F(2,22) = 1.644$, $p = 0.216$).

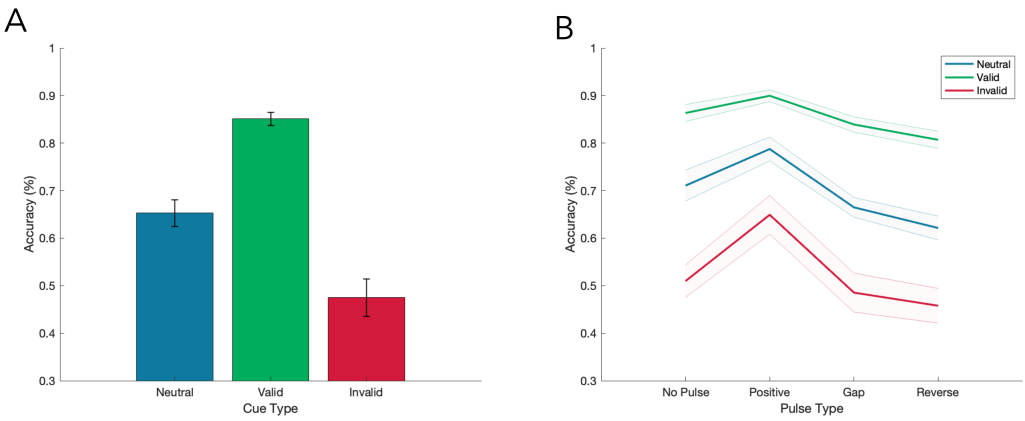


Figure 6.7. *The effect of the cue and pulse on accuracy. A) Overall accuracy in each of the cue conditions. B) Mean accuracy in each pulse and cue condition. The shaded region represents the standard error of the mean. The fact that mean accuracy in the no-pulse, neutral cue condition is ~70% suggested that the difficulty titration worked well.*

Examining how the cue influenced choice accuracy across reaction times offers an insight into how the influence of the prior changes as the trial proceeds (Figure 6.8). A binary logistic mixed effects analysis showed a main effect of Reaction Time ($F(1,62008) = 215.551$, $p < 0.001$) and Cue Condition ($F(2,62008) = 739.773$, $p < 0.001$), and a significant interaction ($F(2,62008) = 216.158$, $p < 0.001$). Accuracy significantly increased on valid cue trials ($\Delta = 16.1 \pm 2\%$, $p < 0.001$) and significantly decreased on invalid cue trials ($\Delta = -18.3 \pm 1.5\%$, $p < 0.001$). The model coefficients were used to characterise how the relationship between cue condition and reaction time influenced accuracy. In the neutral cue condition, accuracy declined as reaction time increased ($\beta = -0.633$, $p < 0.001$). Compared to the neutral cue condition, accuracy declined significantly faster as reaction times increased in the positive pulse condition ($\beta = -0.787$, $p < 0.001$) and accuracy significantly increased across

reaction times in the invalid cue condition (β = 0.97, p < 0.001). This analysis provides some evidence that the effect of the cue erodes as participants spend longer periods of time deliberating.

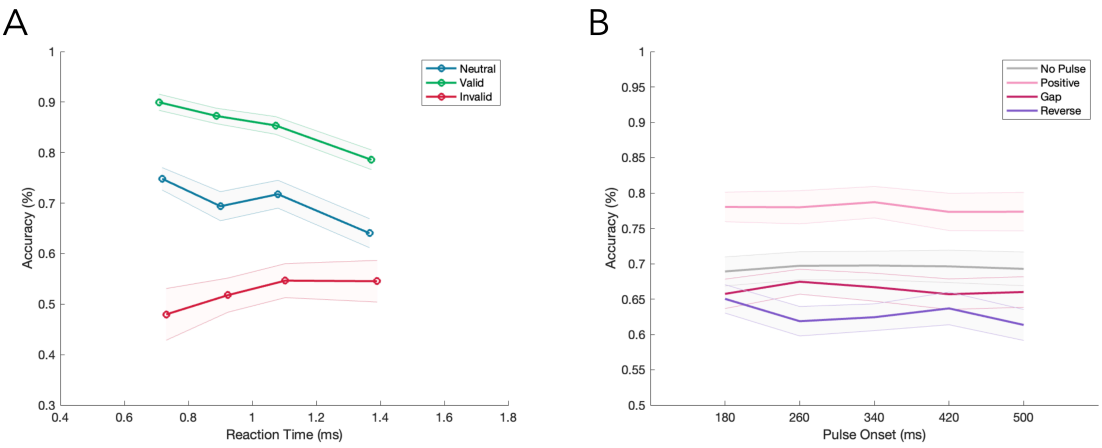


Figure 6.8. *A) The effect of the cue on accuracy across reaction time. The effect of the cue weakens as participants spend longer on the task, but is still present even for the latest reaction time bin. B) The effect of the pulses on accuracy across pulse onset times. On average, there was a small but significant decline in accuracy as pulse onset increased. However, there was no interaction with pulse type.*

To test for a change in the influence of sensory evidence on accuracy as time elapses we applied the same pulse analysis approach as the above reaction time analyses, using the no-pulse condition as a baseline comparison to account for the effect of time on proximity to bound. The influence of the pulse on accuracy at different pulse onsets was analysed with a logistic mixed effects analysis. There was a main effect of Pulse-Type ($F(3,62006) = 70.143$, $p < 0.001$) and Pulse Onset ($F(1,62006) = 13.743$, $p < 0.001$), but no interaction ($F(3,62006) = 1.26$, $p = 0.286$). Compared to the no pulse condition, accuracy was significantly higher on positive pulse trials ($p < 0.001$), and significantly diminished on gap ($p < 0.001$) and reverse pulse trials ($p < 0.001$). The main effect of Pulse Onset was interpreted by repeating the analysis without the interaction term which confirmed that accuracy significantly declined as pulse onset increased ($F(1,62009) = 15.206$, $p < 0.001$).

In summary, the cues and pulses influenced accuracy in the expected directions. Only the effect of the positive pulse appeared to be influenced by the cue condition, with the greatest benefit of positive pulses following invalid cues and the smallest improvement in accuracy for valid cues, suggesting that performance was near a ceiling in this condition. The influence of the cue on accuracy decreased as reaction time increased, but the cue can still be seen to have a clear effect in the slowest reaction time bin (Figure 6.8.A). Finally, there was no evidence that the influence of the pulses on accuracy changed with later pulse onsets.

Electrophysiological Analyses

The analysis of EEG data is presented for each decision-related signal in two phases: before evidence onset and after evidence onset. The interval between cue presentation and evidence onset was examined in order to test for signatures of biases in sensory encoding, evidence accumulation, and motor preparation in the absence of any physical evidence. In the cue-locked analyses, there is no distinction between valid and invalid cues, since no evidence has been presented yet. Therefore, trials from the valid and invalid cue conditions were pooled into a single 'cued' condition (see Methodology). For the SSVEP and MB analyses, the marginal sensory representation/motor preparation favouring the cued stimulus was compared to the marginal sensory representation/motor preparation of the upcoming target in the neutral cue condition. This was chosen as the comparison because there is no way for the participant to forecast the upcoming target, so each stimulus should be equally represented and each response should be equally prepared in the neutral cue condition during the baseline phase.

The Effect of Expectation on Pre-evidence Sensory Encoding

In the following analyses the SSVEP is used to investigate the effects of priors on the neural representation of the sensory evidence feeding the decision-making process. We first sought to replicate previous findings that this signal represents early visual encoding of the physical stimulus and faithfully tracks the contrast of each grating (e.g. Steinemann et al., 2018). As the task required participants to detect the grating with relatively stronger contrast, the sensory evidence was estimated by the 'target marginal SSVEP amplitude' (hereafter 'marginal SSVEP'); that is, the amplitude of the SSVEP signal representing the target grating minus the

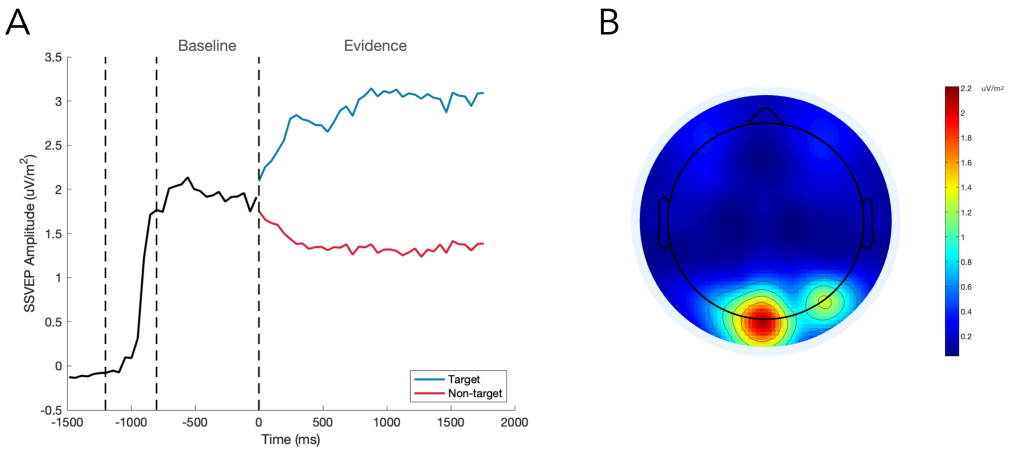


Figure 6.9. *The SSVEP tracks stimulus contrast. A) The dashed time markers represent the onset of the fade-in sequence, the onset of the baseline phase, and the onset of evidence. The mean SSVEP can be seen to rise as the stimulus fades-in. The signal then plateaus for the 800 ms baseline phase, where both gratings are presented at 50% contrast. Finally, the signal clearly discriminates the target stimulus from the non-target stimulus as evidence onsets at 0 ms. B) The topography of the SSVEP signal during evidence presentation shows strong activity over visual cortex.*

amplitude of the SSVEP signal representing the non-target grating. Figure 6.9.A shows that the signal rises as the gratings initially fade-in, plateaus during the baseline phase as the contrast of both gratings is held constant at identical levels, and diverges in opposite directions depending on whether the grating driving the SSVEP increases or decreases in contrast at evidence onset. As expected, the grand average topography of the SSVEP during evidence presentation shows the signal originates in a focal area over occipital cortex (Figure 6.9.B). This neatly demonstrates that the SSVEP tracks stimulus contrast and clearly differentiates the sensory representation of the choice alternatives during evidence presentation.

The first analysis tested the hypothesis that the prior would lead to an enhanced sensory representation of the cued grating in the absence of any physical contrast difference, akin to the preparatory sensory activity discussed in the Introduction (Figure 6.10). A mixed effects model was used to determine if there was a boosting of the SSVEP representation of the cued stimulus in the window from -752:-215 ms[2] (Figure 6.10.B). This was compared to the representation of the upcoming target stimulus in the neutral cue condition. All trials independent of response of response validity were included in this analysis. There was a significant effect of SSVEP Frequency ($F(1,49356.127) = 4119.635$, $p < 0.001$), but no main effect of Cue Condition ($F(1,49356.529) = 1.363$, $p = 0.243$). This result indicates that the cue did not significantly enhance the sensory encoding of the cued stimulus before evidence onsets.

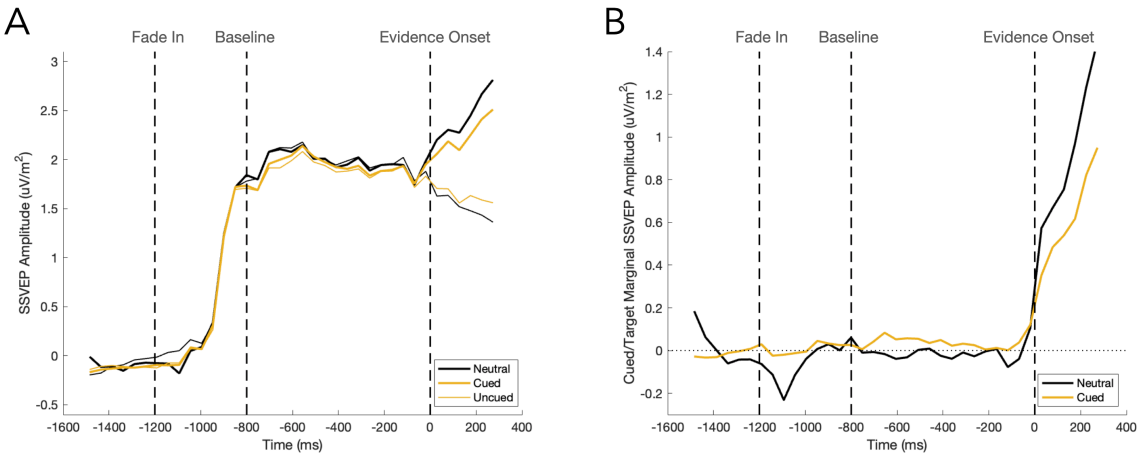


Figure 6.10. *The pre-evidence SSVEP. The time markers represent the onset of the fade-in sequence, the onset of the baseline phase, and the onset of evidence. During the baseline phase, both grating stimuli were presented at 50% contrast. The valid and invalid cue trials were pooled into a 'cued' condition because the validity of the cue is only apparent after evidence has onset. A) The mean amplitude of the SSVEP signals representing each of the stimuli in cued and neutral conditions. In the cued condition, the thick line represents the cued grating and the thin line represents the grating tilted in the opposite direction (uncued). For comparison, in the neutral condition, the thick line is the signal representing the stimulus that will become the target after evidence onset; the thin line is the stimulus that will become the non-target after evidence onset. B) There was no cue effect on the marginal SSVEP amplitude.*

[2] Throughout these analyses the windows chosen for SSVEP and beta analyses may appear oddly specific. This is because the time-frequency transformation produces a timeline with low temporal resolution compared to an ERP like the CPP. This timeline is constructed by taking the mean timepoints in each of the STFT windows sampling the signal. The window reported for these analyses reflects the closest timepoints on the impoverished timeline to the actual window of interest; in this case -750:-200 ms.

The next analysis sought to determine whether the trial-by-trial influence of the pre-evidence SSVEP on behaviour and the subsequent representation of the evidence (demonstrated in Chapter 5) was dependent on the cue (Figure 6.11). Marginal SSVEP refers to the representation of the forthcoming target in all remaining analyses. Since the reaction time and accuracy analyses were dependent on there being a valid response on the trial, only valid responses were included in these analyses. All trials were included in the final analysis as a response was not necessary to have an SSVEP representation of the stimulus.

First, the data were analysed using a mixed effects analysis with SSVEP Target Frequency, Marginal SSVEP, Cue Condition, and a Cue-Marginal SSVEP interaction term to predict reaction time. There was a significant effect of Cue Condition ($F(2,48059.021) = 47.783$, $p < 0.001$), SSVEP Target Frequency ($F(1,48059.033) = 1099.952$, $p < 0.001$), Marginal SSVEP ($F(1,48059.014) = 8.372$, $p = 0.004$), and a significant interaction between Cue Condition and Marginal SSVEP ($F(2,48059.031) = 11.036$, $p < 0.001$). Compared to the neutral cue condition, valid cues reduced reaction time ($p < 0.001$) and invalid cues increased reaction times ($p < 0.001$). This is the same overall effect of cues on reaction time, which was observed in the behavioural analyses. The model's Marginal SSVEP coefficient was negative and significant ($\beta = -0.002$, $p = 0.027$), suggesting that pre-evidence representation of the target leads to faster responses in the neutral cue condition. The effect of pre-evidence representation was equivalent between invalid and neutral cue conditions ($\beta = -0.001$, $p = 0.392$) . The coefficient of the valid cue interaction term was significant ($\beta = 0.004$, $p = 0.002$), indicating that the greater the marginal representation of the target stimulus before evidence onset, the slower the response after evidence onset, relative to the neutral cue condition. The results indicate that having greater pre-evidence sensory representation of the target produces relatively faster responses on trials where the participants have been miscued or received a neutral cue, but once they have been given a valid cue, this effect is reversed and reaction times become relatively slower with greater marginal SSVEP amplitudes (Figure 6.11.B).

The same predictors were used to predict accuracy with a binary logistic regression mixed effects analysis. There was a significant effect of Cue Condition ($F(2,48069) = 2605.003$, $p < 0.001$), SSVEP Target Frequency ($F(1,48069) = 3827.074$, $p < 0.001$), and a significant interaction between Cue Condition and Marginal SSVEP ($F(2,48069) = 4.663$, $p = 0.009$). There was no main effect of Marginal SSVEP ($F(1,48069) = 0.15$, $p = 0.698$). Pairwise comparisons confirmed the result of the behavioural analysis, accuracy was higher on valid cue trials ($p < 0.001$) and lower on invalid cue trials ($p < 0.001$), compared to neutral. Greater pre-evidence representation of the target led to less accurate responses in the neutral cue condition ($\beta = -0.018$, $p = 0.016$). Both the valid cue interaction term ($\beta = 0.018$, $p = 0.04$) and the invalid cue interaction term ($\beta = 0.031$, $p = 0.002$) were positive and significant, indicating that the better the representation of the target in the pre-evidence window in both conditions, the more accurate performance was, compared to the neutral cue condition. While this amounts to an actual improvement in accuracy in the invalid cue condition, the coefficients of the valid and neutral cue interaction terms cancel each other out, meaning that greater pre-evidence marginal SSVEP amplitudes were not associated with any change in accuracy in the valid cue condition (see Figure 6.11.C).

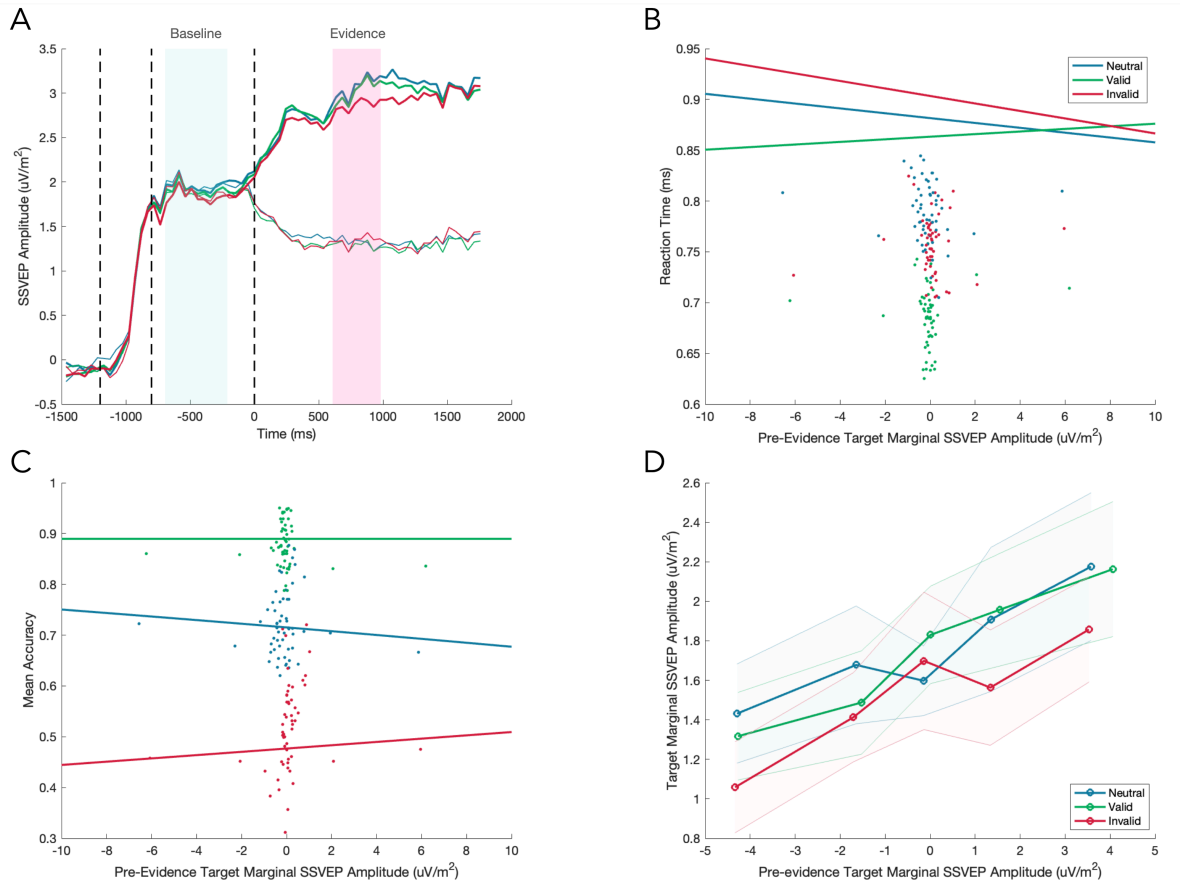


Figure 6.11. *Testing the importance of the sensory representation of the baseline phase of the trial for the subsequent decision. A) Target and non-target SSVEP signals across the full trial, separated by cue condition. The pre-evidence marginal SSVEP amplitude (cyan) was used to predict reaction time, accuracy, and marginal SSVEP during evidence presentation (pink). B & C) Predicting reaction time and accuracy using pre-evidence marginal SSVEP. Each participant's pre-evidence marginal SSVEP amplitudes were placed in ten bins and these binned data points are plotted, coloured by cue condition, alongside the model estimates (solid lines). D) The relationship between pre-evidence marginal SSVEP and marginal SSVEP during evidence presentation. The data were binned by the magnitude of the pre-evidence marginal SSVEP separately for each cue condition. The shaded region represents the standard error of the mean at each bin.*

Finally, the same model was used to predict the marginal SSVEP in the window 684:977 ms during evidence presentation (Figure 6.11.A). This window was chosen to come after the last pulse offset, but not to push too far into the evidence-locked epoch when responses would have been made on most trials. The same window was chosen to analyse the evidence-locked SSVEP in a subsequent section. There was a significant effect of Pre-evidence Marginal SSVEP ($F(1,40109.038) = 305.592$, $p < 0.001$), Cue Condition ($F(2,40109.086) = 5.733$, $p = 0.003$), and SSVEP Target Frequency ($F(1,40109.151) = 1269.625$, $p < 0.001$), but no interaction between the Pre-evidence Marginal SSVEP and Cue Condition ($F(2,40109.174) = 0.739$, $p = 0.478$). The cue effect on the evidence-locked SSVEP will be discussed in a later section, so the main effect was not further investigated here. The analysis was repeated without the interaction term to characterise the main effect of Pre-evidence Marginal SSVEP. Target representation during evidence presentation significantly improved as the pre-evidence marginal SSVEP increased ($\beta = 0.134$, $p < 0.001$).

In summary, these results do not support the hypothesis that the prior augments the representation of the cued stimulus at a sensory level before the evidence onsets, but the sensory representation of the stimulus during this baseline phase was a relevant factor in predicting reaction time, accuracy, and the quality of the subsequent representation of the evidence. This result would be expected if participants began to accumulate evidence prematurely. Alternatively, this could be a reflection of task engagement or attentional state. The finding that accuracy decreases with better pre-evidence representation of the forthcoming target on neutral cue trials is unintuitive. Exploratory analyses were conducted to determine if the change in the sensory representation from pre- to post-evidence was an important factor in the decision process that might help to explain this, but this proved difficult to assess at a single-trial level and ultimately did not provide any conclusive result.

The Effect of Expectation on Pre-evidence Accumulation

Several previous studies have observed early CPP build-up that predicted choice behaviour (Devine et al., 2019; Kelly et al., 2021; McGovern et al., 2018), consistent with a premature accumulation of sensory-noise in anticipation of evidence onset. In this section, the dynamics of the CPP before evidence onset were investigated to determine whether there was pre-evidence accumulation and if so, whether the early CPP build-up was subject to modulation by prior knowledge, consistent with a drift bias or the activation of a sensory template (Feuerriegel et al., 2021b). The increasing amplitude of the CPP during the baseline phase persists long after a visual evoked potential would influence activity (Figure 6.12.A), the topography of activity in this period shows a clear CPP (Figure 6.13.B), and there is a relationship between the CPP amplitude in the baseline phase and subsequent behaviour (discussed below). Together, this suggests that the CPP exhibited substantial build-up prior to the onset of the sensory evidence. A mixed effects analysis was conducted to determine if this

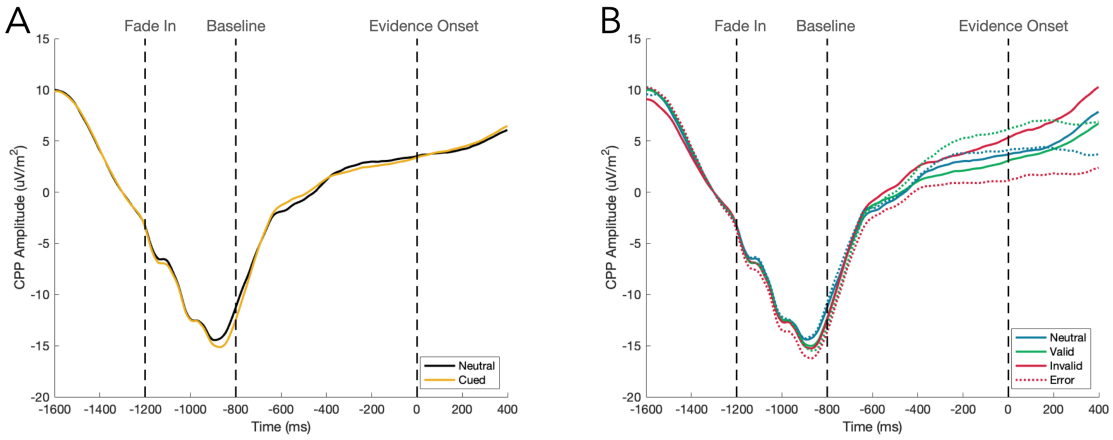


Figure 6.12. *The pre-evidence CPP. A) The pre-evidence CPP shows no difference in amplitude or slope between cued and neutral cue conditions. B) The pre-evidence CPP on correct and error response trials. Increases in the amplitude of the pre-evidence CPP was associated with decreases in accuracy on valid cue trials.*

early build up was influenced by the cue. The CPP was baselined in the window -1400:-1200 ms and its amplitude and slope were measured from -200:0 ms. All trials were included in the cue-locked mixed effects analysis. There was no main effect of Cue Condition on either the amplitude ($F(1,52586.202) = 0.264$, $p = 0.607$) or the slope ($F(1,52586.153) = 0.147$, $p = 0.702$) of the CPP before evidence onset.

This analysis provided no evidence of the cue influencing the CPP before evidence onset. If there was pre-evidence accumulation, which was simply unaffected by the cue, the CPP activity in this baseline period should predict behaviour in the forthcoming trial. To determine if the baseline phase of the trial exerted any influence on the subsequent decision, another set of mixed effects analyses were conducted using the cue condition and CPP amplitude or slope to predict the reaction time and accuracy on the upcoming trial (Figure 6.13). Since these predictions extended past evidence onset, the validity of the cues became relevant, so the analyses used the three cue conditions. In each of these analyses, a main effect of cue condition, simply refers to the behavioural effect on reaction time or accuracy described in an early section, so these effects are reported, but not further explored.

In the analysis predicting reaction time using CPP amplitude, there was a main effect of Cue Condition ($F(2,51239.032) = 53.694$, $p < 0.001$) and a main effect of CPP Amplitude ($F(1,51242.316) = 13.058$, $p < 0.001$), but no interaction ($F(2,51239.028) = 1.099$, $p = 0.333$). The main effect of CPP Amplitude was interpreted by removing the interaction term and running the analysis again to get the overall effect of CPP Amplitude across all cue conditions. The analysis showed that as pre-evidence CPP amplitude increases, the reaction time on the upcoming trial decreases ($F(1,51245.813) = 25.068$, $p < 0.001$; Figure 6.13.C). In the analysis predicting reaction time using CPP slope, there was also a main effect of Cue Condition ($F(2,51239.03) = 55.177$, $p < 0.001$) and a main effect of CPP Slope ($F(1,51243.089) = 3.957$, $p = 0.047$), but no interaction ($F(2,51239.079) = 1.511$, $p = 0.221$). The main effect of slope was characterised by removing the interaction term and repeating the analysis. However, when the interaction term was removed, the slope effect was no longer significant ($F(1,51246.53) = 1.452$, $p = 0.228$; Figure 6.13.D).

In the analysis predicting accuracy using CPP amplitude, there was a main effect of Cue Condition ($F(2,51250) = 2743.797$, $p < 0.001$) and a significant interaction between Cue Condition and CPP Amplitude ($F(2,51250) = 12.427$, $p < 0.001$), but there was no main effect of CPP Amplitude ($F(1,51250) = 0.034$, $p = 0.854$). The interaction was interpreted using the model coefficients. In the valid cue condition, increases in pre-evidence CPP amplitude were associated with a decrease in accuracy ($\beta = -0.001$, $p = 0.007$), compared to the neutral cue condition. There was no significant difference in the relationship between pre-evidence CPP amplitude and accuracy between the invalid and neutral cue conditions ($\beta = 0.001$, $p = 0.088$). Finally, in the analysis predicting accuracy using CPP slope, there was a main effect of Cue Condition ($F(2,51250) = 2697.928$, $p < 0.001$), but no main effect of CPP Slope ($F(1,51250) = 1.952$, $p = 0.162$), and no interaction ($F(2,51250) = 1.149$, $p = 0.317$).

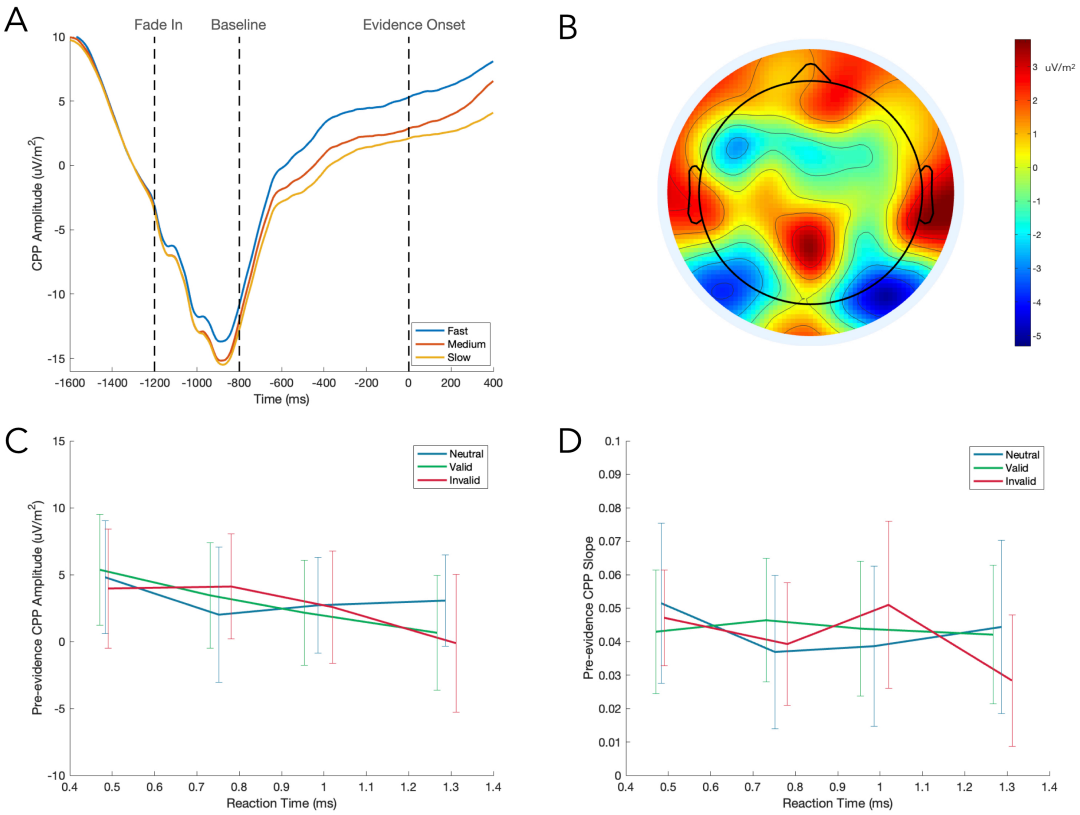


Figure 6.13. *The relationship between pre-evidence accumulation and reaction time. A) The pre-evidence CPP binned by the reaction time on the upcoming trial. B) The topography of the difference between the CPP in the fast and slow reaction time bins during the baseline phase. C) Trials with faster reaction times had greater CPP amplitudes before evidence onset, but this did not change across cue conditions. D) pre-evidence CPP slope does not show a reliable relationship with reaction time on the upcoming trial.*

In summary, these analyses provide evidence that participants started to accumulate evidence during the baseline phase of the trial, but there is no evidence that this early accumulation was subject to an expectational bias. The observation that increased pre-evidence CPP amplitude predicted lower accuracy in the valid cue condition is consistent with the pre-evidence accumulation of noise (e.g. Devine et al., 2019).

The Effect of Expectation on Pre-evidence Motor Preparation

The final component of the pre-evidence decision-making process to be investigated was motor preparation based on the characteristic desynchronisation of contralateral MB activity in anticipation of a manual response. For all MB analyses, more negative amplitudes are indicative of greater motor preparation. To establish if there was preparation of the cued response before evidence onset and determine if the degree of early preparation was related to the subsequent reaction time, a mixed effects analysis was used to predict the amplitude of lateralised MB activity (contralateral minus ipsilateral) in the 200 ms before evidence onset (Figure 6.14.A). All valid responses were included in the analysis. There was a main effect of Cue Condition (F(1,51246.338) =

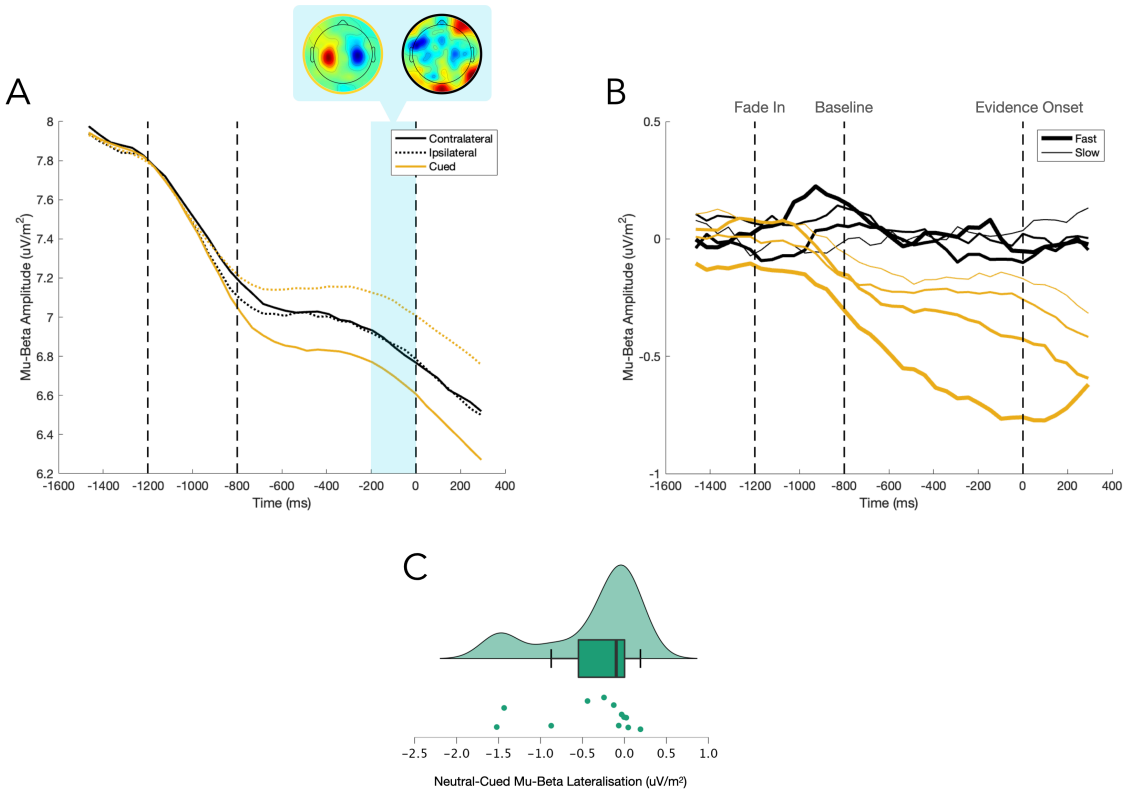


Figure 6.14. *The effect of a cue on motor preparation in cued and neutral conditions. A) Ipsilateral and contralateral mu-beta activity during the baseline phase. The clear lateralisation in the cued condition is illustrated in the difference in the topographies in the window -200:0 ms before evidence onset. The cued condition topography has a yellow surround and the neutral cue condition has a black surround. B) The mu-beta lateralisation by reaction time quartiles. Later reaction time quartiles are shown with thinner lines. This pre-evidence lateralisation in the cued condition is most prominent for trials with fast reaction times, but is substantially reduced on trials where participants took longer to respond. C) The difference in mu-beta lateralisation between neutral and cued conditions for individual participants, shown as a raincloud plot.*

69.479, $p < 0.001$) and Reaction Time ($F(1,51011.424) = 31.541$, $p < 0.001$), and a significant interaction ($F(1,51246.362) = 32.368$, $p < 0.001$). There was significantly greater MB lateralisation in the cued condition compared to the neutral cue condition ($p < 0.001$). To interpret the relationship with reaction time, separate analyses were performed for each cue condition. Faster reaction times were associated with greater pre-evidence MB lateralisation in the cued condition ($F(1,42323.995) = 162.745$, $p < 0.001$), but there was no significant effect of Reaction Time on MB lateralisation in the neutral cue condition ($F(1,8574) = 0.577$, $p = 0.447$). The same factors were used to predict the difference between the slope of contralateral beta and ipsilateral beta in the interval -400:0 ms before evidence onset. There was no main effect of Cue Condition ($F(1,51205.536) = 2.409$, $p = 0.121$) and no interaction ($F(1,51204.897) = 1.075$, $p = 0.3$), but the difference between the slope of ipsilateral and contralateral MB diminished as Reaction Time increased ($F(1,28800.279) = 7.391$, $p = 0.007$). To test for signs that the cue induced a drift bias at the motor-level, the slope of ipsilateral and contralateral signals on cued trials were compared, but there was no significant difference ($F(1,11.315) = 2.398$, $p = 0.149$).

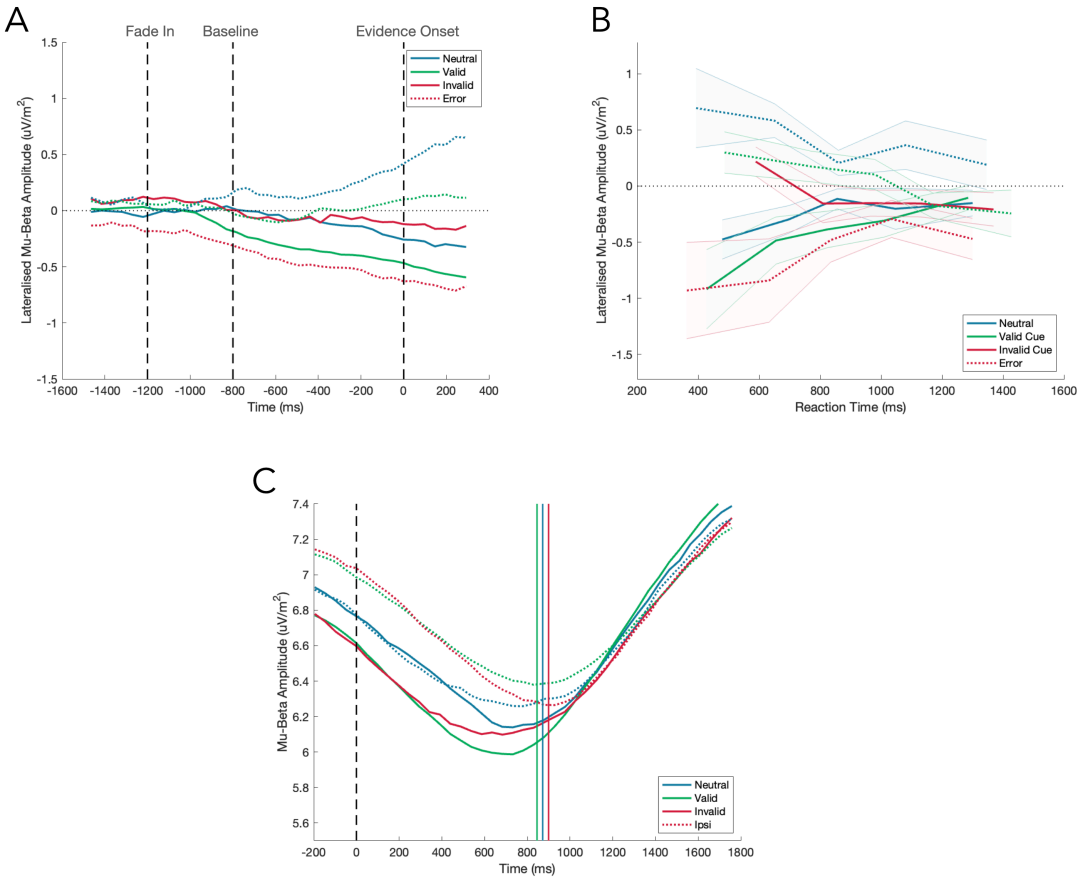


Figure 6.15. *Pre-evidence motor preparation predicts response and reaction time. A) Pre-evidence MB lateralisation for correct and error trials across cue conditions. Here, more negative amplitudes are indicative of preparation of the cued response, or the target response for the neutral cue condition. B) Pre-evidence MB lateralisation binned by subsequent reaction time for correct and error responses across cue conditions. C) Evidence-locked MB plotted relative to the cued response. The influence of pre-evidence lateralisation can clearly be seen to persist after evidence onset. The vertical bars represent the median reaction times in each condition.*

Additional analyses were conducted to replicate previous observations that pre-evidence beta lateralisation is predictive of the forthcoming choice (de Lange et al., 2013), confirming its importance in the decision-making process (Figure 6.15). Again, because the analysis extended into the interval after evidence onset, the three levels of cue validity were included. Two mixed effects analyses were conducted, predicting choice and reaction time using MB lateralisation in the 200 ms before evidence onset. In the analysis predicting choice, there was a main effect of Cue Condition ($F(2,51250) = 2139.978$, $p < 0.001$) and MB Lateralisation ($F(1,51250) = 158.055$, $p < 0.001$), but no interaction ($F(2,51250) = 1.961$, $p = 0.141$). The cue effect on choice accuracy was described in the behavioural analyses. The main effect of MB Lateralisation was interpreted by repeating the analysis with the interaction removed. The probability of a cue-congruent choice increased as pre-evidence MB lateralisation towards that response increased ($F(1,51252) = 164.617$, $p < 0.001$).

In the analysis predicting reaction time for correct responses, there was a main effect of Cue Condition ($F(2,39051.758) = 239.959$, $p < 0.001$), MB Lateralisation ($F(1,39051.755) = 15.093$, $p < 0.001$), and a Cue-Lateralisation interaction ($F(2,39051.196) = 33.748$, $p < 0.001$). In the neutral cue condition, increased MB

lateralisation towards the forthcoming target led to faster reaction times (β = -0.005, $p < 0.001$). Compared to the neutral cue condition, increased preparation of the invalidly-cued response predicted slower reaction times ($p < 0.001$), but there was no significant difference between the neutral and validly-cued trials ($p = 0.056$). For error responses, there was also a main effect of Cue Condition ($F(2,12175.208) = 115.499$, $p < 0.001$), MB Lateralisation ($F(1,12171.249) = 5.105$, $p = 0.024$), and a Cue-Lateralisation interaction ($F(2,12172.727) = 21.774$, $p < 0.001$). Increased MB Lateralisation towards the target predicted slower responses in the neutral cue condition ($p < 0.001$). In comparison, preparation of the invalidly-cued response predicted faster reaction times ($p < 0.001$) and there was no significant difference between valid and neutral cue conditions ($p = 0.529$).

In summary, predictive cues elicit biased motor preparation in advance of evidence onset and the degree of preparation predicts both choice and reaction time. This is not simply reflected in greater contralateral desynchronisation, but also in an equal offset in ipsilateral activity. The relationship between the extent of pre-evidence lateralisation and reaction time provides support for the interpretation of this activity as representing a starting point bias: a smaller starting point bias requires more evidence accumulation to reach the decision bound, and therefore produces longer reaction times (Figure 6.15.B).

The Effect of Expectation on the Encoding of Sensory Evidence

A primary hypothesis in this study was that priors would influence the sensory representation of the stimulus. The analysis of the pre-evidence SSVEP found no significant expectational modulation, but it is conceivable that such a bias could act as a multiplier, enhancing prior-consistent evidence that is present, but not generating an additive boost to the cued stimulus when there was no physical difference between the stimuli. The evidence-locked SSVEP was analysed to address this hypothesis (Figure 6.16). The difference in the marginal SSVEP across cue conditions was investigated with a mixed effects analysis in the evidence-locked window 684:977 ms. This window was chosen to assess the representation of the sensory stimulus without the disruption of the pulses, as the final pulse ends at 650 ms. There was a main effect of Cue Condition ($F(2,45266.07) = 4.221$, $p = 0.015$) and SSVEP Target Frequency ($F(1,45266.177) = 1992.17$, $p < 0.001$). Compared to neutral cue trials, the SSVEP representation of the target was significantly reduced on invalidly-cued trials ($p = 0.036$), but there was no significant difference on valid cue trials ($p = 0.809$; see Fig 6.16.A).

It appears there is an effect of the cue on the sensory representation of the stimulus during evidence presentation. However, it is possible that this effect could be driven by responses, since the median reaction time is marginally faster in the valid cue condition than the invalid cue condition. For example, if there was a stereotyped transient boost in marginal SSVEP amplitude following a response, independent of the cue condition, the reaction time difference could allow this to masquerade as a genuine sensory level modulation. To ensure the effect was not accounted for by differences in reaction times across cue conditions, a mixed effects analysis was conducted on the marginal SSVEP in the response-locked window -200:0 ms (see Figure 6.16.C). Like the evidence-locked analysis, this model used cue condition and SSVEP target frequency as

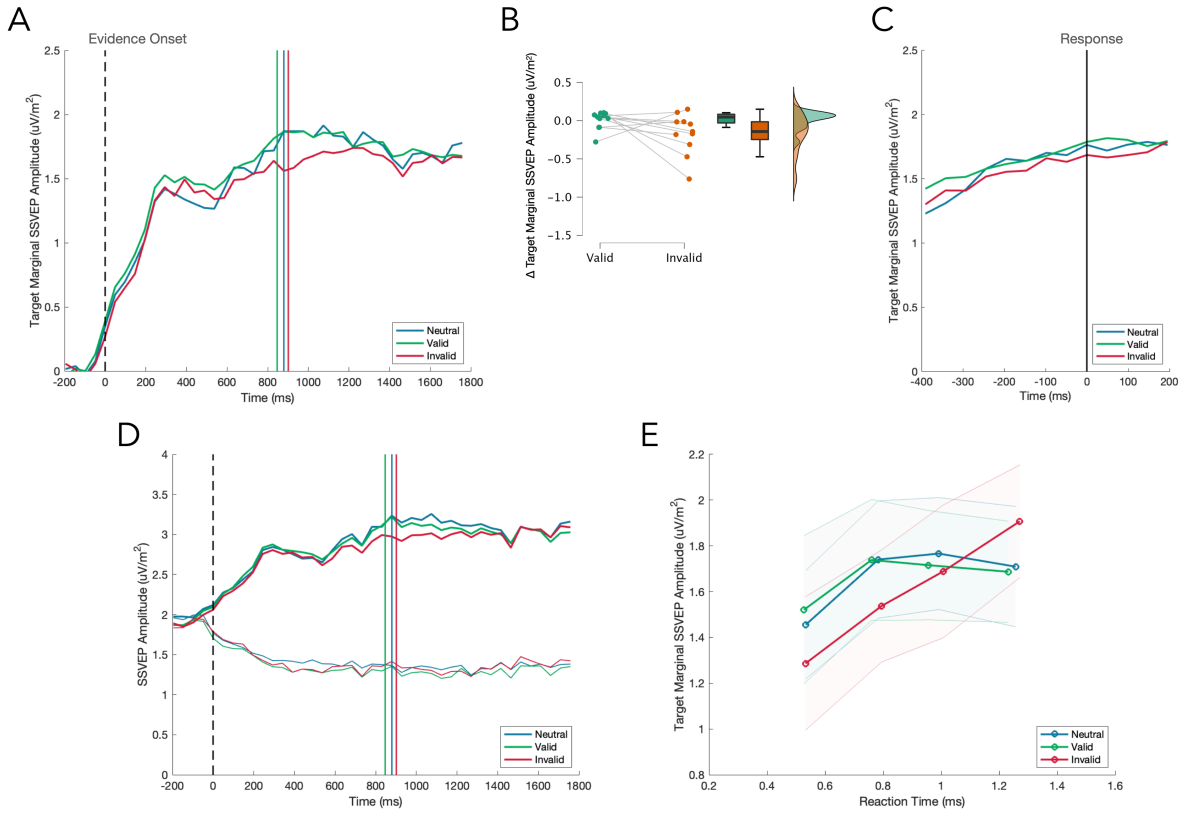


Figure 6.16. *The effect of a cue on sensory representation during evidence presentation. The vertical markers show the median reaction time for the associated cue condition. A) The marginal SSVEP amplitude for each of the cue conditions across evidence presentation. B) Individual differences in marginal SSVEP amplitude between neutral and each of the cued conditions, presented as a raincloud plot. C) The response-locked marginal SSVEP amplitude for each cue condition. D) The mean amplitude of the SSVEP signals representing the target (thick line) and non-target (thin line) for each cue condition. E) The marginal SSVEP amplitude at response shown in reaction time quartiles.*

predictors. To account for differences in the target representation depending on how early or late in the trial the response was made, reaction time was included as a predictor. It was also possible that any cue effect would become stronger/weaker as reaction time increased, so an interaction between reaction time and cue condition was also included. There was a main effect of Cue Condition ($F(2,42672.242) = 4.778$, $p = 0.008$), Target SSVEP Frequency ($F(1,42673.1) = 2215.108$, $p < 0.001$), Reaction Time ($F(1,42654.686) = 32.861$, $p < 0.001$), and a significant interaction between Reaction Time and Cue Condition ($F(2,42672.304) = 3.944$, $p = 0.019$). Neither valid ($p = 0.766$) nor invalid cue trials ($p = 0.354$) differed significantly from the neutral cue condition. However, the model coefficients indicated that this was attributable to the significant interaction between reaction time and cue condition[3] (see Figure 6.16.E). To provide more detail about this interaction, separate analyses were conducted for each cue type. There was a significant increase in marginal SSVEP at response as Reaction Time increased in the valid cue condition ($F(1,28314.681) = 11.774$, $p < 0.001$) and invalid cue condition

[3] When a continuous variable, in this case reaction time, is included in a mixed effects analysis, the model coefficients for the other categorical factors are based on the situation where that continuous variable is at its minimum value. The pairwise comparisons take a different approach, comparing the main effect of a categorical variable when the continuous variable is at its median value. In this analysis, the model coefficients indicated that the difference between the neutral and invalid cue conditions approached significance ($p = 0.056$) when reaction time was set to zero, but this p-value had increased substantially when reaction time was set to the median value of 883 ms. Based on this, it was concluded that the cue effect on the marginal SSVEP may be obscured by the interaction with reaction time.

(F(1,7024.743) = 22.777, p < 0.001), but there was no significant main effect of Reaction Time on marginal SSVEP in the neutral cue condition (F(1,7063.455) = 2.632, p = 0.105).

Overall, these results provide evidence for a small expectational modulation of the sensory representation of the stimulus, which appears to weaken as reaction times extend further into the trial. One of the primary examples of a sensory modulation discussed in the literature is repetition suppression, where repeated exposure to a stimulus evokes a reduced response compared to a stimulus alternation (Grill-Spector et al., 2006). A follow-up analysis was conducted to determine if this SSVEP effect was purely driven by the cue or if this effect was modulated according to whether the target was the same as the target on the previous trial or the opposite target from previous trial. Initial inspection of the repetition-alternation difference waveform indicated that there were two phases that distinguished the SSVEP response on a repetition trial from an alternation (Figure 6.17.C).

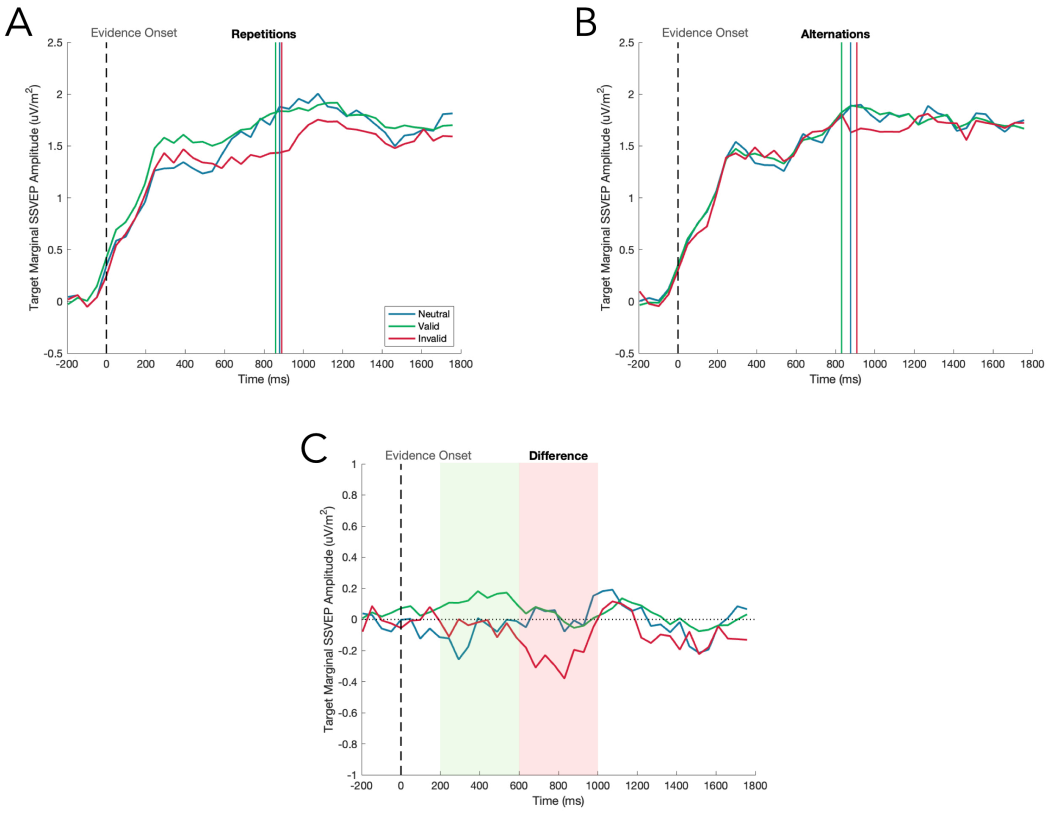


Figure 6.17. *The effect of a cue on marginal SSVEP amplitude on target repetition (A) and alternation trials (B). The vertical markers show the median reaction time for the associated cue condition. C) The effect of repetition on the cue effects is illustrated in the difference (repetition minus alternation) waveform. The early and late phases used in the analyses are highlighted.*

In the early phase (200:600 ms), there was a significant main effect of Cue Condition (F(2,44120.071) = 3.186, p = 0.041), SSVEP Target Frequency (F(1,44120.196) = 3135.281, p < 0.001), and a significant interaction between Cue Condition and Target Repetition/Alternation (F(2,44120.096) = 3.098, p = 0.045). There was no main effect of Target Repetition/Alternation (F(1,44120.138) = 0.133, p = 0.715). The marginal SSVEP was

significantly greater in the valid cue condition compared to the neutral cue condition (p = 0.016). There was no significant difference between invalid and neutral cue trials (p = 0.351). The interaction was investigated with separate analyses for each cue type. Marginal SSVEP amplitude was significantly greater on target repetition trials in the valid cue condition (F(1,29308.194) = 6.345, p = 0.012). There was no significant difference between target repetition and target alternation trials in the neutral cue (F(1,7420.509) = 0.609, p = 0.435) or invalid cue condition (F(1,7370.806) = 0.981, p = 0.322).

In the later phase (600:1000 ms), there was a significant main effect of Cue Condition (F(2,44120.055) = 4.291, p = 0.014) and SSVEP Target Frequency (F(1,44120.152) = 2021.125, p < 0.001), but no interaction between Cue Condition and Target Repetition/Alternation (F(2,44120.075) = 2.936, p = 0.053) and no main effect of Target Repetition/Alternation (F(1,44120.107) = 2.334, p = 0.127). Pairwise comparisons showed that marginal SSVEP amplitude was significantly reduced in the invalid cue condition compared to the neutral cue condition (p = 0.02), but there was no significant difference between the valid and neutral cue conditions (p = 0.933).

If the cue produced a sensory-level modulation, it was expected that this effect would be exaggerated in the response to pulses of sensory evidence (Figure 6.18). A mixed effects analysis of the marginal SSVEP in the pulse-locked window 0:250 ms was conducted. The marginal SSVEP was re-baselined in the window -250:-50 ms to isolate the pulse effect. There was a significant main effect of Pulse-Type (F(3,42396.536) = 223.063, p < 0.001), but no main effect of Cue Condition (F(2,42397.557) = 0.507, p = 0.602) and no interaction between

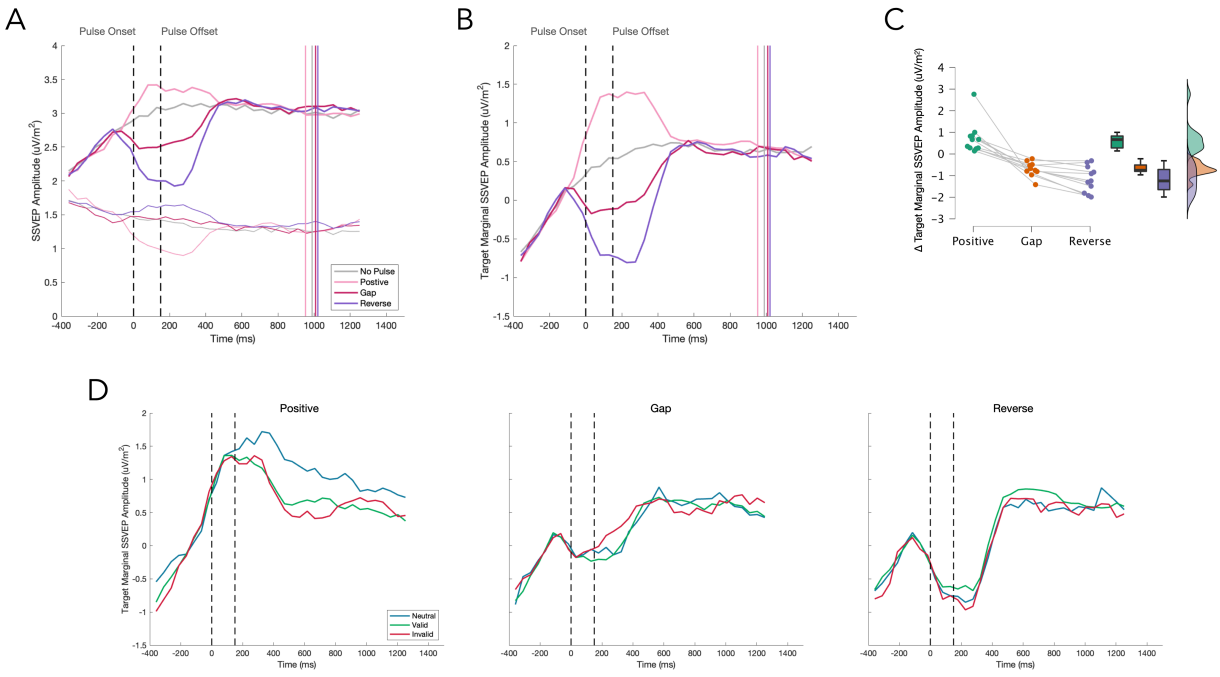


Figure 6.18. *SSVEP response to pulses of sensory evidence. The plots are pulse-locked; the time markers indicate the onset and offset of the pulse. The vertical lines mark the median reaction time in each pulse condition. A) The response of the target (thick line) and non-target (thin line) SSVEP signals to each of the pulse types. B) The pulse-locked target marginal SSVEP amplitude. C) Individual differences in marginal SSVEP amplitude response to the pulse, relative to the no pulse condition, shown as a raincloud plot. D) The pulse-locked target marginal SSVEP amplitude for each cue condition.*

Pulse-Type and Cue Condition ($F(6,42395.941) = 0.722$, $p = 0.631$). There was also no main effect of SSVEP Target Frequency ($F(1,42401.34) = 0.051$, $p = 0.821$). Positive pulses significantly increased the marginal SSVEP ($p < 0.001$), gap pulses significantly reduced the marginal SSVEP ($p < 0.001$), and reverse pulses reduced the marginal SSVEP significantly more than a gap pulse ($p < 0.001$).

This analysis establishes that the SSVEP was sufficiently sensitive to the sensory evidence to register the pulses, but that the encoding of the pulses was not significantly influenced by the cue preceding their presentation. This conflicts with the earlier analysis of the evidence-locked SSVEP suggesting that there was a significant expectation-based modulation of the sensory representation of the evidence. It was possible that the influence of the prior changes dynamically across the duration of the trial; indeed, the previous analysis suggested that the cue effect was primarily found amongst earlier reaction times. In this case, the impact of the pulse would change across different pulse onsets, depending on its congruence with the cue. For example, if the influence of a sensory prior decreased across the duration of the trial, one would expect that on a valid cue trial, a late positive pulse would have a smaller impact on the marginal SSVEP than an early positive pulse. Conversely, the response to an early reverse pulse may be enervated by this sensory drift bias, but grow increasingly influential as the prior effect dissipates. To test this hypothesis, the sign of data on gap and reverse pulse trials, which have already been shown to cause a significant reduction in marginal SSVEP, was reversed so that the analysis was carried out on the magnitude of the effect without directionality.

There was a main effect of Pulse-Type ($F(3,42392.668) = 138.221$, $p < 0.001$), but no main effect of Pulse Onset ($F(1,42396.441) = 0.365$, $p = 0.546$) and no main effect of Cue Condition ($F(2,42393.325) = 0.105$, $p = 0.9$). There was a significant interaction between Pulse-Type and Pulse Onset ($F(3,42392.515) = 71.866$, $p < 0.001$), but no Cue-Pulse interaction ($F(6,42392.494) = 0.791$, $p = 0.577$). The interaction was characterised with separate mixed effects analyses for each pulse-type. The effect of a positive pulse significantly decreased as its onset time increased ($F(1,10564.922) = 44.968$, $p < 0.001$). Conversely, the effect of a gap ($F(1,10722.382) = 63.531$, $p < 0.001$) and reverse pulses ($F(1,10718.772) = 55.392$, $p < 0.001$) significantly increased across onset times.

This result suggests that pulses in the opposite direction of the evidence on each trial become increasingly influential on the sensory representation of the stimulus as their onset time increases. However, after careful examination, it was concluded that this result was a predictable feature of the SSVEP signal. As can be seen in Figure 6.19.A, at all onset times the sliced no-pulse marginal SSVEP signal is still rising. It is not clear to what extent this reflects the underlying sensory encoding of the contrast stimulus and how much of this is accounted for by the temporal smearing of the 400 ms STFT window. Regardless, it appears that the marginal SSVEP takes some period of time to reach a stable amplitude representing the sensory evidence (see sliced no-pulse data in Figure 6.19.B). The later in this rising phase that a positive pulse onsets, the smaller the marginal SSVEP amplitude increase required to represent the increased contrast. Similarly, the later the onset of a gap or reverse pulse, the larger the decrease in marginal SSVEP amplitude needed to represent the absence or reversal of evidence (see Figure 6.19.C). Therefore, it was concluded that the apparent change in the impact of

a pulse across onset times likely reflected the nature of the signal being examined and could not be said to reflect the reality of the underlying sensory processing.

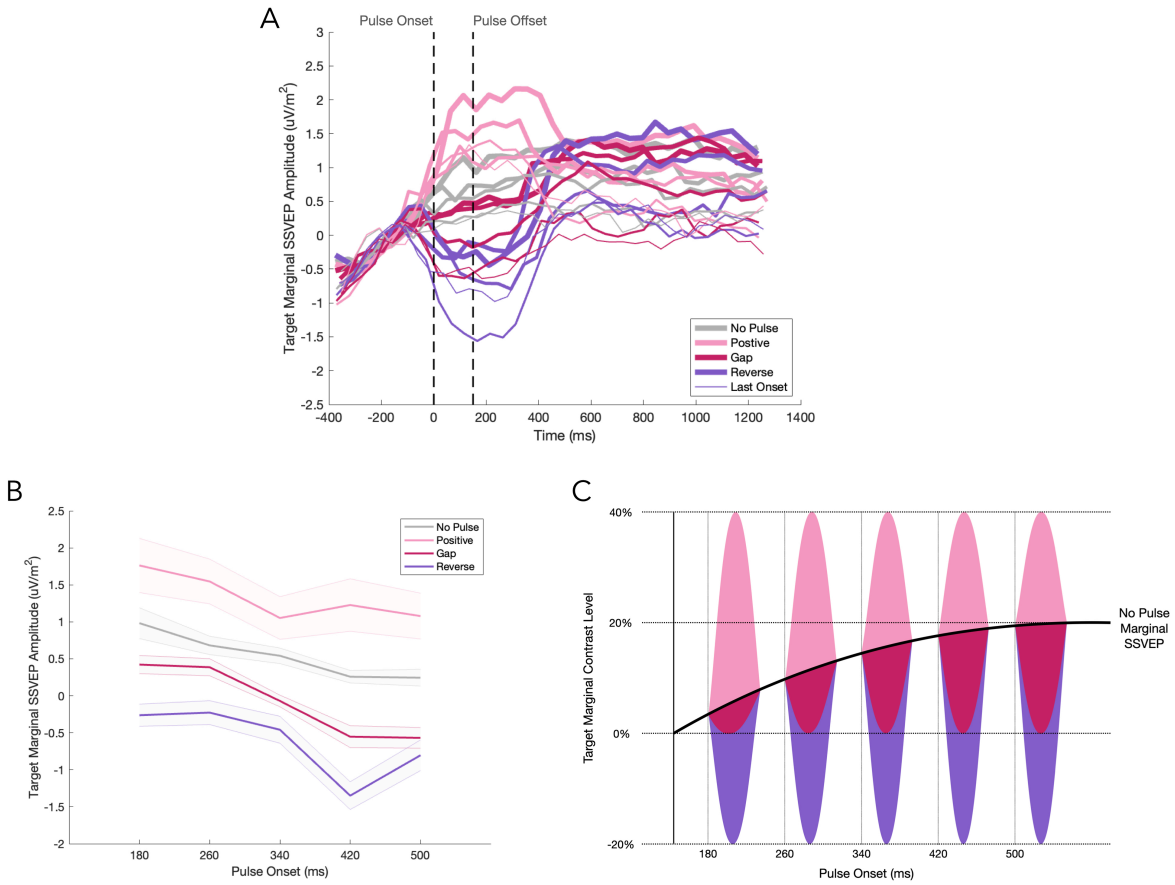


Figure 6.19. *SSVEP response to pulses at different pulse onsets. A) The plot is pulse-locked; the time markers indicate the onset and offset of the pulse. The marginal SSVEP is baselined in the window -250:-50 ms. The thickest line is the earliest onset and the thinnest line is the latest onset. Based on this plot it appears that positive pulses become less influential as pulse onset increases, while gap and reverse pulses become increasingly influential at later onsets. However, this can be explained by the nature of the SSVEP signal itself. B) The baselined target marginal SSVEP amplitude in the window 0:250 ms for each pulse type across pulse onset times. The declining amplitude in the sliced no-pulse condition indicates that, even when there are no pulses, the marginal SSVEP is still rising in this period of the trial. C) An illustration of the explanation for the apparent change in pulse effects across onset times. In this example, the standard contrast evidence on the trial is 60% for the target and 40% for the non-target (i.e. 20% marginal contrast). The thick black line represents the marginal SSVEP in the no-pulse condition, which rises over an extended period spanning the pulse onsets. At later onsets, the increase in marginal contrast required to represent the contrast level during a positive pulse gets smaller as the marginal SSVEP climbs. Conversely, the decrease in the marginal SSVEP needed to reflect the gap or reversal of evidence increases as the marginal SSVEP grows across pulse onsets.*

Finally, to further establish that the SSVEP was an adequate approximation of the sensory evidence accumulated in the decision-making process, the marginal SSVEP was used to predict the reaction time and accuracy on a trial-by-trial basis using the marginal SSVEP amplitude (684:977 ms after evidence onset; Figure 6.20). Only trials where the response occurred after 680 ms were included. A linear mixed effects analysis was used to analyse the reaction time data and a binary logistic regression mixed effects analysis was used to analyse the accuracy data. When predicting reaction time, there was a main effect of Marginal SSVEP

(F(31230.243) = 7.175, p = 0.007), SSVEP Target Frequency (F(1,31226.166) = 450.137, p < 0.001), and Cue Condition (F(2,31221.432) = 34.645, p < 0.001), but no interaction between Marginal SSVEP and Cue Condition (F(1,31221.197) = 0.957, p = 0.384). The main effect of Marginal SSVEP was characterised by repeating the analysis and removing the interaction term. Reaction time significantly decreased as marginal SSVEP increased (β = -0.0014, p < 0.001). When predicting accuracy, there was a main effect of Marginal SSVEP (F(1,31231) = 8.819, p = 0.003), SSVEP Target Frequency (F(1,31231) = 1708.975, p < 0.001), and Cue Condition (F(2,31231) = 1055.558, p < 0.001), but no interaction between Marginal SSVEP and Cue Condition (F(2,31231) = 0.42, p = 0.657). Again, the main effect of Marginal SSVEP was characterised by repeating the analysis and removing the interaction term. Accuracy significantly increased as marginal SSVEP increased (β = 0.013, p < 0.001).

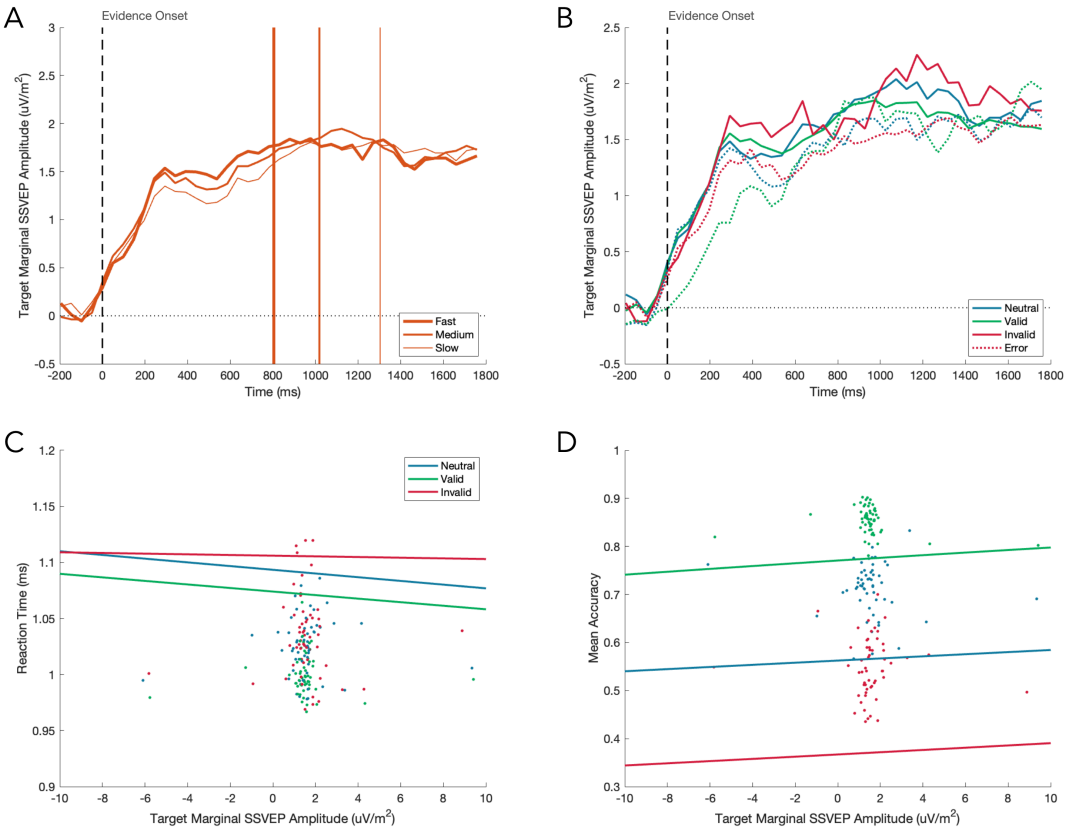


Figure 6.20. *Target marginal SSVEP amplitude during evidence presentation significantly predicts reaction time and accuracy. Like the analyses, only trials with reaction times greater than 680 ms were included in these plots. A) The marginal SSVEP binned by trials with fast, medium, and slow reaction times. The vertical lines indicate the median reaction time in each of the three bins. B) The marginal SSVEP for correct and error trials, separated by cue condition. C & D) Using the marginal SSVEP to predict reaction time and accuracy, respectively. Each participant's data was placed in ten bins according to the magnitude of the marginal SSVEP amplitude. The binned data points, coloured by cue condition, are shown alongside the model parameters (solid lines).*

In summary, the representation of the sensory evidence was found to be modulated by cue condition during evidence presentation, such that the marginal SSVEP amplitude was reduced on invalidly-cued trials. A response-locked analysis indicated that this effect was not explained by different reaction times across the cue conditions. Furthermore, the effect seems to be primarily found on repetition trials exhibiting two distinct

phases of cue modulation; the first showing increased marginal SSVEP for valid cue trials and the second showing reduced marginal SSVEP for invalid cue trials. In contrast, expectation did not exert an influence on the SSVEP responses to the pulse. It could not be determined if this was due to the strength of the expectation effect changing across the course of the trial. Finally the validity of the SSVEP as an index of the encoding of sensory evidence was supported by the observations that evidence-locked marginal SSVEP amplitude predicted reaction time and accuracy.

The Effect of Expectation on Evidence Accumulation

The earlier analysis of the CPP found no evidence of biased accumulation before evidence onset, but there was a cue effect on the sensory information feeding the decision process so it is plausible that the drift bias primarily or exclusively exerts an influence once evidence is being presented. If this was the case, one would expect that accumulation in the direction of the cue would lead to more rapid ascents to the bound, and therefore the slope of the CPP at response would be steeper when the response was consistent with the cue. Alternatively, if the effect of the prior was better accounted for by a starting point shift, one would expect that the amplitude of the CPP at response would be greater when the participant's choice contradicted the cue (Figure 6.21.B).

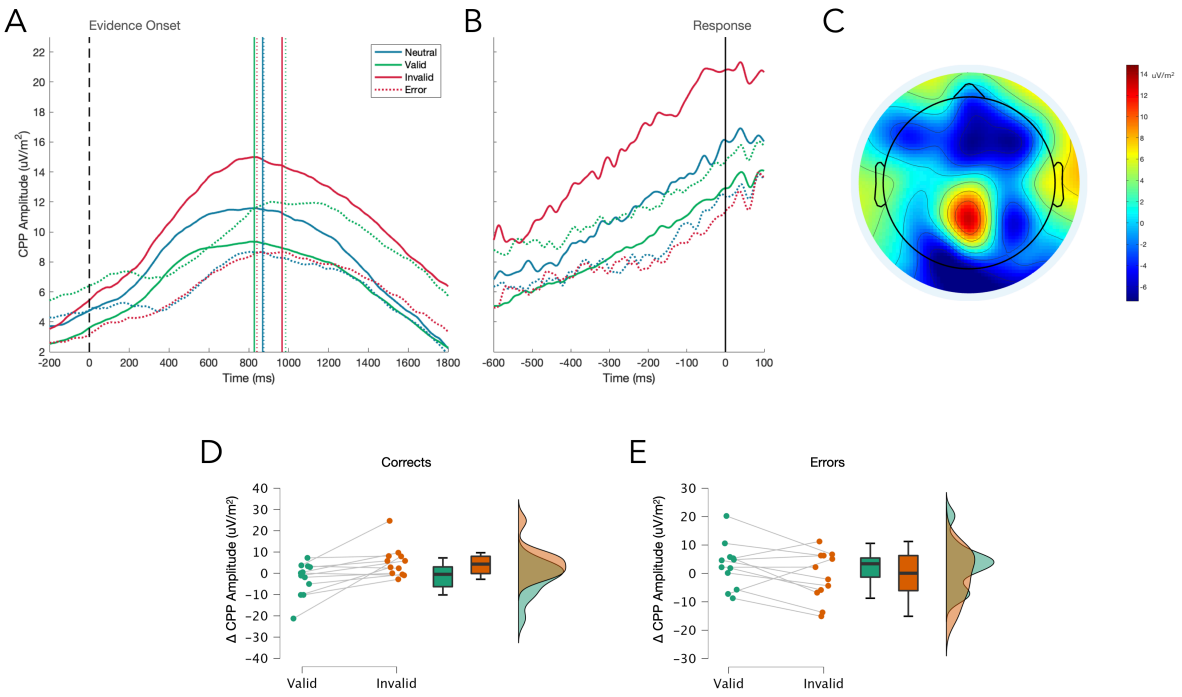


Figure 6.21. *The cue effect on the CPP for correct and error trials. A) The evidence-locked CPP. The vertical lines indicate the median reaction time in the associated conditions. B) The response-locked CPP. C) The response-locked topography, showing the CPP. D & E) Individual differences in CPP amplitude for cued relative to neutral cue trials, shown as separate raincloud plots for corrects (D) and errors (E).*

Mixed effects analyses were conducted on the CPP slope and amplitude in the 200 ms window prior to the response, using trials with response times greater than 200 ms. There was no main effect of Cue Condition ($F(2,45703.818) = 1.699$, $p = 0.183$) or Accuracy ($F(1,45711.766) = 0.068$, $p = 0.794$) on the CPP slope at response, and there was no interaction between Cue Condition and Accuracy ($F(2,45707.652) = 1.809$, $p = 0.164$). In the analysis of the CPP amplitude, there was a main effect of Accuracy ($F(1,45683.595) = 14.024$, $p < 0.001$) and a significant interaction between Accuracy and Cue Condition ($F(2,45659.890) = 14.154$, $p < 0.001$), but no main effect of Cue Condition ($F(2,45708.127) = 1.29$, $p = 0.275$). CPP amplitude was significantly greater on correct response trials ($p < 0.001$). The interaction was interpreted with separate analyses of correct and error trials.

For correct response trials, there was a main effect of Cue Condition ($F(2,34766.371) = 13.67$, $p < 0.001$). The CPP amplitude was significantly greater on invalid cue trials compared to neutral cue trials ($p < 0.001$), but there was no significant difference between valid and neutral cue trials ($p = 0.082$). For error trials, there was also a main effect of Cue Condition ($F(2,10260.255) = 4.036$, $p < 0.018$). Compared to neutral cue errors, there was no significant difference in the CPP amplitude on either valid ($p = 0.075$) or invalid cue errors ($p = 0.509$). However, CPP amplitude was significantly greater on valid compared to invalid cue errors ($p = 0.006$). A final analysis assessed whether there may have been an initial difference in the CPP trajectory across cue conditions that had lapsed by the time of the response. The slope was measured from 0:400 ms and trials with response times shorter than 400 ms were excluded. There was no main effect of Cue Condition on the slope of the early CPP ($F(2,43400.112) = 0.327$, $p = 0.721$).

These analyses do not support the drift bias hypothesis. Instead, the CPP amplitude difference across cue conditions is consistent with a starting point bias. However, the pulses were included in the paradigm specifically to test for drift bias in evidence accumulation, so a second mixed effects analysis was conducted on the pulse-locked CPP to confirm this result (Figure 6.22). If the drift bias accelerated the accumulation of cue-consistent evidence, the CPP response to a pulse of evidence in the same direction as the cue should lead to a much larger response than the response to the same pulse of evidence when the cue was neutral or pointed in the opposite direction. Only trials with response times at least 250 ms after pulse onset were included in the pulse-locked analyses. The CPP amplitude was baselined in the window -50:50 ms and the CPP slope was baselined in the window -200:0 ms, relative to pulse onset. The response to the pulse was measured using the slope of the CPP in the 200 ms after pulse offset and the amplitude of the CPP from 100:400 ms after pulse offset.

In the analysis of the CPP amplitude, there was a main effect of Pulse-Type ($F(3,45170.297) = 17.979$, $p < 0.001$) and a significant interaction between Pulse-Type and Accuracy ($F(3,45172.627) = 10.947$, $p < 0.001$). There was no main effect of Cue Condition ($F(2,45176.612) = 1.879$, $p = 0.153$) or Accuracy ($F(1,45085.947) = 2.762$, $p = 0.097$), and no interaction between Cue Condition and Accuracy ($F(2,45095.04) = 2.111$, $p = 0.121$). Crucially, there was no interaction between Pulse-Type and Cue Condition ($F(6,45169.075) = 0.718$, $p = 0.635$) and no three-way interaction ($F(6,45170.112) = 1.51$, $p = 0.17$). Compared to the no-pulse condition, CPP amplitude

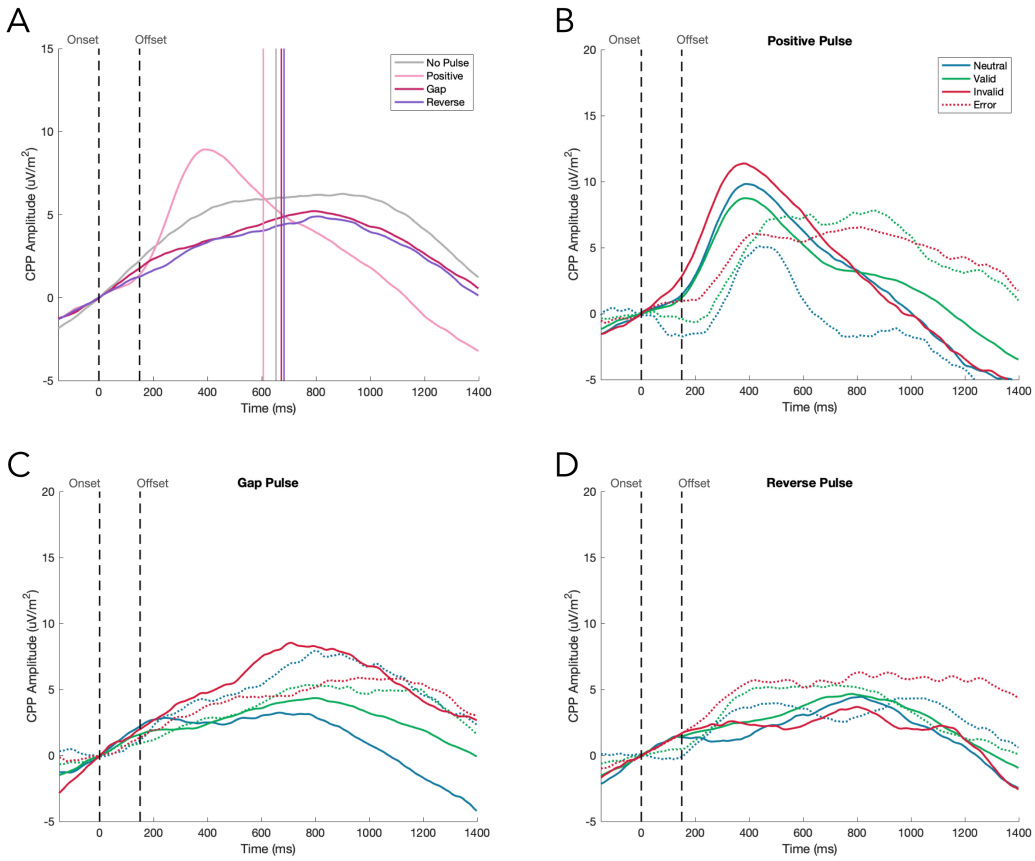


Figure 6.22. *A) The effect of the evidence pulse on the CPP. The vertical lines indicate the median reaction times in each pulse condition. The pulse-locked CPP on correct and error trials across cue conditions for B) the positive pulse, C) the gap pulse, and D) the reverse pulse. The results suggest that a drift bias does not shape the CPP response to the pulses.*

was significantly greater following a positive pulse ($p = 0.003$) and significantly reduced following a gap pulse ($p = 0.001$) or a reverse pulse ($p < 0.001$). The relationship between pulse-type and accuracy was investigated in Chapter 5, so the interaction is not explored here.

In the analysis of the pulse-locked CPP slope, there was a main effect of Pulse-Type ($F(3,45175.876) = 33.557$, $p < 0.001$) and Accuracy ($F(1,42441.635) = 4.877$, $p = 0.027$), there was also a significant interaction between Pulse-Type and Accuracy ($F(3,45172.583) = 24.02$, $p < 0.001$). There was no main effect of Cue Condition ($F(2,45130.281) = 0.05$, $p = 0.951$) and no interaction between Cue Condition and Accuracy ($F(2,43146.035) = 2.264$, $p = 0.104$). Importantly, there was also no interaction between Cue and Pulse-Type ($F(6,45172.659) = 0.717$, $p = 0.636$), and no three-way interaction ($F(6,45175.365) = 1.111$, $p = 0.353$). Compared to the no-pulse condition, the CPP slope was significantly steeper following a positive pulse ($p < 0.001$), but there was no significant difference in CPP slope following a reverse pulse ($p = 0.569$) or a gap pulse ($p = 0.851$). The pulse-locked CPP slope was also significantly steeper on error response trials ($p = 0.027$). The relationship between pulse-type and accuracy was explored in Chapter 5.

These analyses also fail to provide any evidence for a prior-induced drift bias at the evidence accumulation stage of the decision-making process. The final drift bias hypothesis was that the influence of a drift bias may change across the duration of the trial (Hanks et al., 2011; Deneve, 2012), increasingly divorcing evidence accumulation from the physical evidence at later periods of the trial. This dynamic drift bias would be expected to augment the impact of cue-congruent pulses and dampen the impact of cue-incongruent pulses at later onsets on trials where the response matched the cue. However, this may be complicated for error trials by the fact that the evidence points in the opposite direction to the response for most of the trial, which could potentially influence the drift bias. For this reason, the analysis used only correct response trials. Although it was unlikely, it was possible that this change in the influence of a drift bias across the trial obscured the effect in the previous analyses that pooled across pulse onsets, so a set of mixed effects analyses were conducted on the CPP amplitude and slope, including pulse onset as a predictor (Figure 6.23).

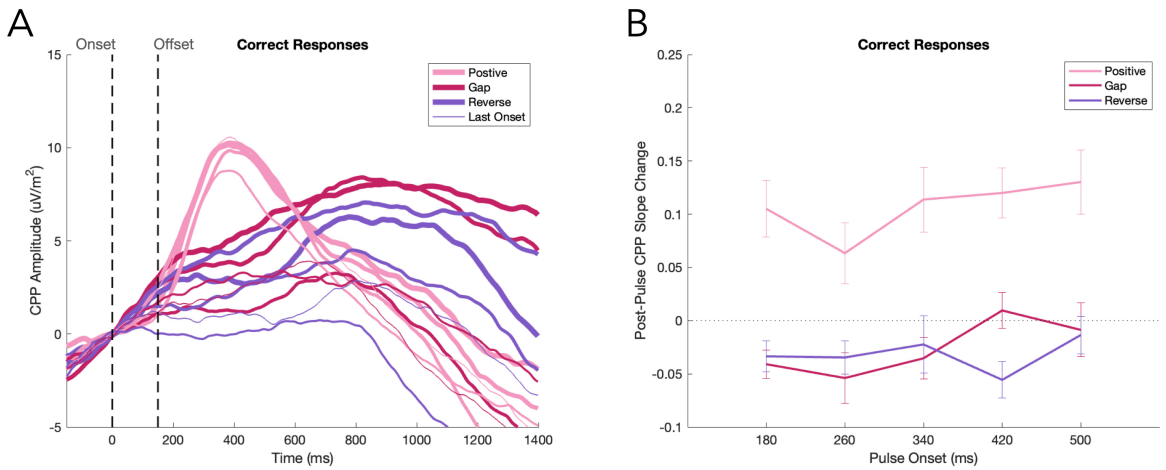


Figure 6.23. *The effect of the pulse on the CPP across pulse onsets, for correct response trials. A) The CPP is shown for each pulse onset, where thicker lines indicate earlier onsets. B) The pulse effect on the slope of the CPP is shown across pulse onsets. The data have been baselined by the no pulse condition. The error bars represent the standard error of the mean.*

The only significant result of the post-pulse CPP amplitude analysis was a main effect of Pulse-Type ($F(3,34358.269) = 11.886$, $p < 0.001$). There was no significant main effect of Cue Condition ($F(2,34362.607) = 1.21$, $p = 0.298$) or Pulse Onset ($F(1,34359.604) = 0.062$, $p = 0.803$). There was also no significant interaction between Cue Condition and Pulse-Type ($F(6,34358.073) = 0.473$, $p = 0.829$), or between Cue Condition and Pulse Onset ($F(2,34358.184) = 0.142$, $p = 0.868$), or between Pulse-Type and Pulse Onset ($F(3,34357.596) = 0.855$, $p = 0.464$), and no three-way interaction ($F(6,34357.559) = 0.234$, $p = 0.966$). Compared to the no-pulse condition, CPP amplitude was significantly greater following a positive pulse ($p < 0.001$) and significantly reduced following a gap pulse ($p < 0.001$) or a reverse pulse ($p < 0.001$).

In the analysis of post-pulse CPP slope, there was a main effect of Pulse-Type ($F(3,34362.965) = 32.837$, $p < 0.001$) and Pulse Onset ($F(1,34367.689) = 7.252$, $p = 0.007$). There was no main effect of Cue Condition

(F(2,34335.444) = 0.345, p = 0.708), and no interaction between Pulse-Type and Cue Condition (F(6,34361.815) = 1.017, p = 0.412), Cue Condition and Pulse Onset (F(2,34362.405) = 0.047, p = 0.954), or Pulse-Type and Pulse Onset (F(3,34360.475) = 0.734, p = 0.532). There was also no three-way interaction (F(6,34359.847) = 0.621, p = 0.714). Compared to the no pulse condition, CPP slope was significantly greater following a positive pulse (p < 0.001) and significantly reduced following a gap pulse (p = 0.024) or reverse pulse (p < 0.001). The main effect of onset time was interpreted by repeating the analysis, but removing all interaction terms. As onset time increased, the CPP slope was increasingly positive after a pulse, independent of the pulse type (F(1,34371.582) = 14.636, p < 0.001).

Finally, several previous studies have shown that the CPP's amplitude at response can vary as a function of reaction time due to the influence of urgency processes at the motor-level pushing motor preparation signals closer to their threshold as time elapses (Kelly et al., 2021; Steinemann et al., 2018). In their investigation of the effects of expectation in perceptual decision-making, Kelly et al. showed that urgency can lead to complicated expressions of expectation-based model adjustments in response-locked CPP amplitudes, which they could only interpret with careful modelling. To determine if urgency may also be a complicating factor in the manifestation of expectation modulations in the current dataset, the amplitude of the CPP in the 200 ms preceding the response was compared across cue conditions and reaction times in separate mixed effects analyses for correct and error responses (Figure 6.24).

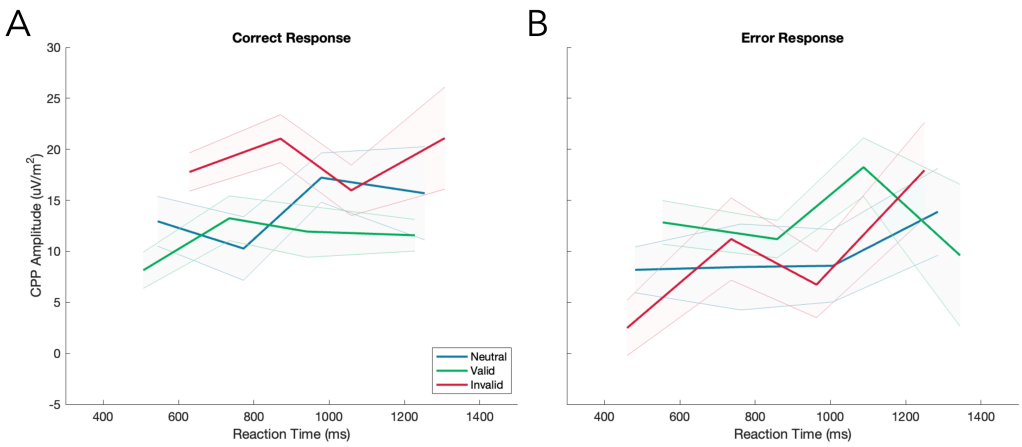


Figure 6.24. *The amplitude of the CPP at response across cue conditions and reaction time quartiles, for (A) correct and (B) error response trials. The shaded areas represent the standard error of the mean.*

For correct responses, there was a main effect of Reaction Time (F(1,33857.144) = 3.845, p = 0.05), but no main effect of Cue Condition (F(2,34766.829) = 0.76, p = 0.468), and no interaction (F(2,34766.087) = 0.459, p = 0.632). The Reaction Time effect was interpreted by removing the interaction term and repeating the analysis. CPP amplitude increased as Reaction Time increased (F(1,31053.778) = 6.049, p = 0.014). For error responses, there was also a main effect of Reaction Time (F(1,8285.177) = 21.247, p < 0.001), but no main effect of Cue Condition (F(2,10897.443) = 0.8, p = 0.449) and no interaction (F(2,10931.435) = 0.181, p = 0.835). The main

effect of Reaction Time was interpreted the same way and this analysis produced the same result, CPP amplitude increased as Reaction Time increased ($F(1,8150.935) = 21.257$, $p < 0.001$).

This was an unexpected and perplexing result, but it was anticipated that these results would be difficult to interpret without modelling (see Figure 6.2.B). However, as a representation of accumulated evidence, collapsing bounds would be expected to reduce the CPP amplitude for later reaction times in the neutral cue condition, where there should be no starting point bias and no drift bias. So the analyses were repeated using only the neutral cue condition. There was no effect of Reaction Time on CPP amplitude at response for correct responses ($F(1,4121.881) = 0.921$, $p = 0.337$), but it was still found to increase as reaction time increased for error responses ($F(1,658.713) = 4.88$, $p = 0.027$).

In summary, there was no evidence of a drift bias in the cue effects on the response-locked CPP or in the CPP response to pulses under different cue conditions, and there was no evidence that the failure to detect a drift bias could be attributed to the influence of the bias changing across the duration of the trial. Instead, the response-locked CPP results indicate that the best explanation for the effect of priors at the evidence accumulation level is a starting point shift. Attempts to identify hallmarks of urgency in the CPP produced ambiguous results.

The Effect of Expectation on Post-Evidence Motor Preparation

It has already been established that the cue prompted greater preparation of the cued response before evidence had onset, but it is possible that this preparatory offset was quickly overwhelmed by the urgency and evidence signals that drive MB to its threshold during evidence presentation. It was hypothesised that if pre-evidence motor preparation had a lasting influence on the decision process as it evolved across evidence presentation, there would be greater desynchronisation in the MB signal representing the unchosen response (ipsilateral) when the response contradicted the cue. Additionally, in this scenario, the chosen response would need to overcome the preparatory offset to reach the threshold at response and this accelerated preparation should manifest as a steeper contralateral MB slope (Figure 6.25).

The influence of the cue on motor preparation was initially measured in the slope of contralateral MB in the 300 ms before response with a mixed effect analysis. There was a significant interaction between Cue Condition and Accuracy ($F(2,45704.882) = 19.627$, $p < 0.001$), but no main effect of Cue Condition ($F(2,45702.585) = 0.609$, $p = 0.544$) or Accuracy ($F(1,45705.983) = 2.849$, $p = 0.091$). The interaction was interpreted with separate analyses for correct and error responses. For correct responses, there was a main effect of Cue Condition ($F(2,34762.58) = 15.431$, $p < 0.001$). There was steeper contralateral MB desynchronisation on invalid cue trials than neutral cue trials ($p < 0.001$), but there was no significant difference between valid and neutral cue conditions ($p = 0.055$). For error responses, there was also a significant main effect of Cue Condition ($F(2,10935.95) = 8.742$, $p < 0.001$). The slope of contralateral MB was significantly steeper on valid cue trials

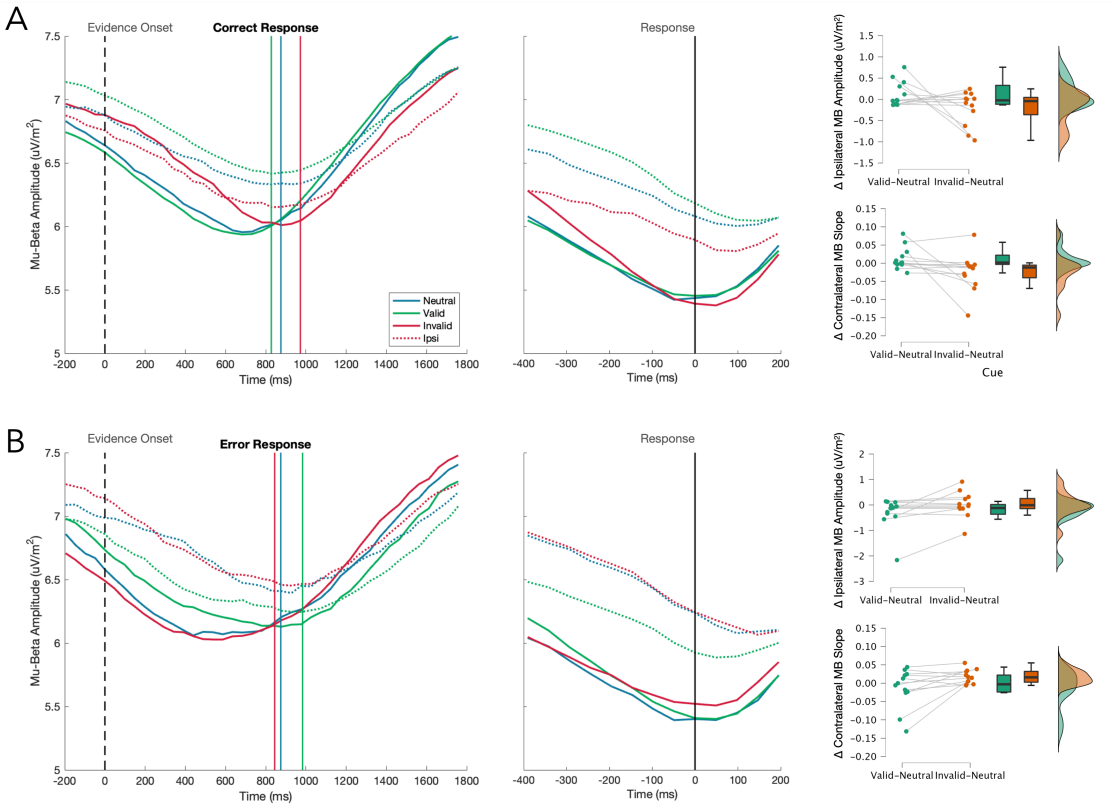


Figure 6.25. *The effect of the cue on motor preparation in evidence-locked and response-locked plots. The plots show ipsilateral and contralateral beta-band activity for correct (A) and error (B) response trials. Ipsi- and contralateral refer to the response. The vertical lines represent the median reaction time in the associated condition. The individual differences between valid and neutral and between invalid and neutral cue conditions are shown in raincloud plots for the two significant results: contralateral MB slope and ipsilateral MB amplitude.*

compared to neutral cue trials (p = 0.033), but there was no significant difference between invalid and neutral cue conditions (p = 0.121).

Next, the amplitude of ipsilateral mu-beta in the 100 ms before the response was compared across conditions with a mixed effects model using the same predictors as the slope analysis. There was a main effect of Accuracy ($F(1,45702.621) = 8.052$, $p = 0.005$) and a significant interaction between Cue Condition and Accuracy ($F(2,45702.472) = 30.15$, $p < 0.001$), but no main effect of Cue Condition ($F(2,45702.084) = 2.011$, $p = 0.134$). Ipsilateral MB desynchronisation was significantly increased on correct response trials (p = 0.005). Separate analyses of correct and error responses were conducted to characterise the interaction. There was a main effect of Cue Condition on correct response trials ($F(2,34761.233) = 19.096$, $p < 0.001$). Ipsilateral MB exhibited significantly less desynchronisation on valid cue trials ($p < 0.001$) and significantly more desynchronisation on invalid cue trials (p = 0.02), compared to neutral cue trials. There was also a main effect of Cue Condition on error response trials ($F(2,10931.422) = 12.876$, $p < 0.001$). Ipsilateral MB exhibited significantly greater desynchronisation on valid cue trials ($p < 0.001$), but there was no significant difference between invalid and neutral cue trials (p = 0.862).

In summary, the cue-induced bias in pre-evidence motor preparation appears to have a lasting influence on the decision process throughout evidence presentation. When the cue pointed in the opposite direction to the eventual response, motor preparation for the cued alternative was significantly more advanced at response than when a neutral cue had been presented. Conversely, when the response was consistent with the cue, ipsilateral desynchronisation was significantly less advanced. Given that after evidence onset the trials are identical across cue conditions, it seems very likely that the beta effects observed at response are in part the legacy of the pre-evidence adjustments. The persistence of this pre-evidence bias would demand that responses which contradict the cue must undergo more rapid contralateral MB desynchronisation to overcome this deficit. Indeed the slope of contralateral MB was significantly steeper on correct response trials when an invalid cue had been presented and on error response trials when a valid cue had been presented.

It had been planned to explore the impact of the pulses on motor preparation in different cue contexts, but the pulse effects on MB were too small to be analysed. The temporal smearing associated with the 400 ms STFT window is a likely culprit in masking the effects, since the pulses only lasted 150 ms. The lateralised readiness potential, which was analysed in Chapter 5 as an index of motor preparation, also was not amenable to analysis because the reduction in trial count when pulse trials were segregated by cue condition and response (left/right) rendered the signal too noisy to reliably detect the pulse responses.

Discussion

Predictive cues have repeatedly been shown to have a substantial influence on behaviour in perceptual decision-making tasks (Mulder et al., 2012). The Drift Diffusion model provides an algorithmic description of the decision process underlying these effects, but it is agnostic as to whether these effects are implemented at the level of sensory encoding, evidence accumulation, or motor preparation, so neurophysiological data are needed to characterise the mappings from behaviour to mechanism (Gold & Stocker, 2017). This study investigated the two candidate implementations of expectation-based modulations provided by sequential sampling models with analyses of participant behaviour and electrophysiological indices of three key components of the decision process in a cued contrast discrimination task.

Expectation and Sensory Encoding

In the past decade, a substantial literature has developed reporting a variety of expectation-based sensory modulations. Some of these studies suggest that expectation leads to the generation of anticipatory feature-specific sensory templates, which shape the processing of the subsequent stimulus or may themselves be accumulated as early evidence (Feuerriegel et al., 2021b). In this study, efforts were made to capture any expectational modulations of sensory activity before or during evidence presentation by tracing the dynamics of the SSVEP. However, there was no evidence that the cue elicited preferential encoding of the expected grating orientation during the baseline phase of the trial. There are several caveats to this result: most of the studies reporting preparatory sensory activity either applied decoding analyses to neurophysiological recordings (e.g. Kok et al., 2017) or directly measured neural activity in specific sensory areas like area MT (e.g. Schlack & Albright, 2007); the preparatory activity has been reported to be significantly weaker than the responses to physical stimuli; and none of these studies identified evidence of preparatory activity with an SSVEP. For these reasons, it is also a possibility that the SSVEP analysis was not sufficiently sensitive to detect the anticipatory activity. On the other hand, the SSVEP was sufficiently sensitive to capture modulations in the representation of the sensory evidence in this baseline period that predicted reaction time, accuracy, and the quality of the subsequent representation of the physical contrast evidence. However, it is not clear whether this indicates that this early sensory representation was actually fed into the decision process or if this result stems from changes in trial-to-trial task engagement.

In contrast to the pre-evidence analysis, cues did significantly impact on the marginal amplitude of the SSVEP following evidence onset, with significantly smaller responses following invalid cues. The discrepancy between this result and the SSVEP result reported in Chapter 3 is discussed in Chapter 7. A common observation in the expectation literature is that evoked responses to unexpected stimuli tend to be larger than responses to

expected stimuli (Feuerriegel et al., 2021a). This effect has been associated with predictive processing schemes, where unexpected sensory events provoke prediction error signals that call for a revision of perceptual hypotheses (Bell et al., 2016; Egner et. al, 2010; Heilbron & Chait, 2017). However, the SSVEP exhibited the opposite response to expectation, with a smaller response to the unexpected stimulus.

One account of how expectations influence sensory processing suggests that the responses in neurons tuned to the expected stimulus features are enhanced, while those tuned to non-expected features are suppressed (de Lange et al., 2018). Since the preponderance of the population will be tuned to non-expected features, this results in a global suppression, but that suppression arises from a sharpening of the representation of the expected stimulus (e.g. Kok et al., 2012). The SSVEP result is consistent with this account because the neural activity captured in each SSVEP signal is necessarily driven by units tuned to each stimulus, since only these units would be expected to respond at the specific frequency of the stimulus modulation eliciting the SSVEP. This is an important advantage of the SSVEP analysis over analyses of BOLD activity that are restricted to describing the effects of expectation based on the average stimulus preferences of voxels, which contain intermingled conglomerations of differently-tuned neurons. The result, therefore, could be explained by the invalid cue enhancing responses to the grating with the weaker contrast and weakening responses to the grating with the stronger contrast. The relative parity of these SSVEP signals, compared to the neutral cue condition, would have produced a smaller marginal SSVEP signal overall. It would also be expected that a valid cue should have the converse effect, such that the SSVEP signal representing the expected grating should be enhanced and the SSVEP signal representing the unexpected grating should be dampened, producing a stronger marginal SSVEP overall. However, the failure to detect this may be attributable to a ceiling effect. It is possible that the baseline contrast difference was large enough that a valid cue did little to further strengthen the representation of the sensory evidence, but there was sufficient scope to weaken the evidence by boosting the representation of the competing grating on invalid cue trials. Although there was no consequent reduction in the rate of evidence accumulation on invalid cue trials, it is not entirely clear that a small sensory modulation of this nature would be readily detectable in the slope of the CPP amidst competing influences on evidence accumulation. This might be clarified by comparing the CPP time course to model-simulations of this effect.

The paradigm was specifically designed to test for sensory-level modulations by amplifying the potential effects of biased encoding using brief pulses of evidence. If sensory responses to expected stimuli were enhanced, a cue-congruent pulse should provoke stronger responses than the same pulses in the absence of any expectation. Although the pulses were clearly identifiable in the SSVEP signals, there was no evidence that the responses were systematically modulated by the cue-context in which they were presented. Previous studies have suggested that the macroscopic response to an expected stimulus may be deceptive because it may actually be driven by the divergent adjustments to the response properties of sensory populations based on how their feature preferences align with the expectation (Kok et al., 2012; Summerfield & de Lange, 2014). However, the converse tuning adjustments associated with sharpened representation cannot explain the indifference of the pulse response to the cue because, as mentioned above, the SSVEP is generated by neurons-tuned to the stimulus.

The dissociation of expectation and attention has been a Gordian obstacle in expectation research, particularly due to the prevalence of cueing paradigms (Aitchison & Lengyel, 2017; Rungratsameetaweemana & Serences, 2019; Summerfield & Egner, 2009). The use of a cuing paradigm in this study means that the sensory effect cannot be definitively attributed to sensory priors. In fact, feature-based attention provides an explanation for the apparent inversion of the expectation suppression: if an invalid cue led participants to pay closer attention to the non-target orientation, this may have degraded the differential representation of the contrast evidence. It is possible that the reason attention did not significantly enhance the marginal SSVEP when valid cues were presented was because the physical evidence had already guided attention to the target grating on neutral and valid cue trials. However, the pulse responses would also be expected to be amplified by attention and the unexpected finding that the sensory effect was primarily evident on trials where the target was repeated from the previous trial further complicates this interpretation. The response on repetition trials appears to differ from the response on alternation trials in two distinct phases. In an early phase, the validly-cued representation of the target is enhanced, while there is no significant difference between invalid and neutral cue trials; in the later phase, the invalidly cued representation of the target is impoverished, while there is equivalent target representation in valid and neutral cue trials. It is not obvious why feature-based attention would be guided by the cue only on trials where the target repeats. One may also question what role attention could play in this task. Participants were required to determine which grating was presented with relatively higher contrast, so both gratings were always task-relevant and it is not clear that preferentially processing the cued grating would aid performance (Summerfield & Egner, 2016). This could potentially be addressed in a follow up study comparing psychophysical thresholds and SSVEP amplitudes on a first block where participants are explicitly told to attend to one grating orientation, but not told that it is more likely to be the target, to a second block where they receive an explicit probabilistic cue.

Another commonly reported effect in the expectation literature is that repeated stimuli tend to elicit dampened responses compared to alternated stimuli, a phenomenon termed repetition suppression (Grill-Spector et al., 2006). The literature on the potential relationship between expectation and repetition suppression is very much unsettled. Many conflicting observations have been reported (Bell et al., 2017; Feuerriegel et al., 2021a; Vinken & Vogels, 2017) and there is debate about whether repetition suppression is a low-level version of expectation suppression instantiating a basic temporal prior or if it is an entirely distinct processing mechanism associated with adaptation (Auksztulewicz & Friston, 2016; Grotheer & Kovacs, 2016). However, there have been reports that expectation suppression interacts with repetition suppression (Larsson & Smith, 2012; Summerfield et al., 2011; Utzerath et al., 2017). For example, Summerfield et al. (2011) also found an expectation-repetition interaction in ERPs evoked over central electrodes by paired face stimuli. They found that the difference between the response to expected and unexpected stimuli was greater for repetition trials than alternation trials, 300:400 ms after stimulus onset. However, Feuerriegel et al. (2018) also measured ERPs in response to briefly presented face pairings, which were independently expected/unexpected and repetitions/alternates. They found multi-phased and topographically distributed expectation and repetition effects, but no interaction between expectation and repetition.

The sensory modulation by priors reported in the present chapter is a perplexing observation because it does not neatly accord with the characteristics of either an expectational or an attention-based effect and the use of a cueing paradigm does not allow for these accounts to be disentangled. While there are some observations of interactions, many studies have failed to find any relationship between repetition suppression and expectation (Feuerriegel et al., 2018, 2021a; Kovacs et al., 2013; Kaliukhovich & Vogels, 2011) and others indicate that the relationship can be completely changed when attention is also manipulated (Larsson & Smith, 2012; Richter & de Lange, 2019). The relationship between sensory signals (e.g. ERPs or BOLD activity) examined in these studies and the decision process is uncertain, but the SSVEP provides a selective representation of the sensory evidence feeding the decision. The lack of any prior modulation in the response to the pulses does raise questions about the effect, but it has been previously reported that the SSVEP response to changes in contrast can be non-linear (Kim et al., 2007); it is possible the SSVEP response saturated at the higher contrast levels associated with the positive pulse and there is a small trend consistent with the overall cue effects on the SSVEP in the reverse pulse condition (Figure 6.18.D). Although it is unclear whether this effect was driven by expectation, attention, or both, the result was a diminished representation of the contrast evidence on invalidly-cued trials. This may be a candidate mechanism for a sensory-level driven drift bias. However, further research is needed to clarify the origins of this effect.

The Decisions Mechanisms Responsible For Implementing Expectation

The preponderance of investigations of the role of expectation in perceptual decision-making have endorsed an explanation based on a preparatory shift in the starting point of evidence accumulation towards the boundary associated with the expected choice alternative. The primary objective of this study was to determine if this account sufficiently describes the underlying neurophysiological implementation.

The behavioural data showed the characteristic effects of expectation: reaction times were faster for cue-congruent decisions and slower for cue-incongruent decisions, valid cues increased accuracy and invalid cues decreased accuracy. One of the key differences in the starting point bias and drift bias accounts is the predictions they make about reaction time distributions; a starting point bias produces a sharp change in the leading edge of the cued distribution, but a drift bias account predicts a pervasive change across the entire distribution (Diederich & Busemeyer, 2006). While there was an increase in early cue-congruent responses, the differences across cue conditions in both correct and error distributions were still evident for the latest reaction times. However, the behavioural influence of the pulses did not provide any support for the drift bias account. Although the pulses significantly influenced reaction time, their influence was not contingent on the cue. There was an interaction between cues and pulses in the effects on accuracy: a positive pulse had a bigger effect on accuracy in the invalid cue condition. However, this interaction may result from a ceiling effect. The mean accuracy in the valid cue condition when there was no pulse was greater than 85%, this does not leave much

scope for improvement when a positive pulse is introduced. In contrast, mean accuracy in the no-pulse invalid cue condition was approximately 50%, so additional evidence would almost certainly have a substantial impact.

Independent of the pulse, the cue effect on accuracy is also ambiguous. A starting point bias should have a substantial, but transient impact on choice accuracy that is no longer evident for the longest reaction times. Conversely, a drift bias should persist as decision time elapses. The effect of the cue did diminish as reaction time increased, but the cue effect persists across all reaction time bins. It might be argued that the contingent of trials with long reaction times may be disproportionately composed of trials where the participants ignored the cue. However, as can be seen in Figure 6.8.A, even in the final quartile of responses, there is still a substantial difference in accuracy across the cue conditions, so it is highly unlikely that the participants did not register the cue on these trials. This suggests that the prior is a particularly potent influence on the evolution of the early decision process, but with greater exposure to the evidence, the relative influence of the prior and sensory evidence is increasingly balanced.

The cue did not influence the amplitude of the CPP before evidence onset, consistent with previous reports (Kelly et al., 2021; Steinemann et al., 2018). This result was not surprising as the CPP exhibits the characteristics of a pure, motor-independent representation of the cumulative evidence (O'Connell & Kelly, 2021). However, it has been suggested that preparatory sensory activity, resembling anticipatory templates of the expected stimulus, may serve as an early source of evidence for the decision (Feuerriegel et al., 2021b), which would be expected to increase the amplitude of the CPP before evidence onset. Although this is a compelling hypothesis, the failure to detect any evidence of this with either the SSVEP or the CPP strongly suggests that it did not occur in this study. The pre-evidence CPP amplitude was a significant predictor of subsequent reaction time, which indicates that participants may have begun to accumulate before evidence onset (as observed by Kelly et al., 2021). In the absence of any cue influence on this process, and given that the physical stimuli are presented at equal contrast, it appears that this was simply the premature accumulation of noise. It is therefore not surprising that greater pre-evidence accumulation predicted lower accuracy in the valid cue condition, when simply following the cue would lead to 80% accuracy.

Although the cue did not influence the pre-evidence CPP, there was a clear bias in motor preparation favouring the cued response as indexed by premotor mu-beta signals. This boost in preparatory activity for the cued alternative emerged shortly after the cue was presented and was sustained up to evidence onset. Replicating de Lange et al. (2013), this activity was shown to be decision-relevant; the magnitude of the motor-level bias predicted the timing and accuracy of the forthcoming choice. Indeed, these biases persisted throughout decision formation and the legacy of the starting point offsets was readily detectable in the amplitude of ipsilateral and the slope of contralateral mu-beta immediately preceding the response. Together, this indicates that the cue evoked a starting point bias, instantiated in preparatory motor activity, favouring the more probable perceptual hypothesis.

At the evidence accumulation stage, a drift bias would be expected to accelerate evidence accumulation for cue-congruent responses, which was expected to be evident in an increase in the slope of the response-locked CPP (Figure 6.21.B). However, the CPP at response exhibited no change in slope across cue conditions and a follow-up analysis showed no difference in CPP slope across cue conditions in the first 400 ms of evidence presentation either. Rather the cue effect manifested in differentiated CPP amplitudes at response, with significantly greater amplitudes for correct responses on invalid cue trials and error responses on valid cue trials. The cue effect on CPP amplitude is neatly accounted for by a starting point bias, as both correct responses on invalid cue trials and error responses on valid cue trials require the decision variable to overcome an initial deficit in order to reach the corresponding threshold. This account was further supported by the observation that, like the SSVEP, neither the CPP's amplitude nor its slope were prejudiced by the cues in responding to the pulses. The pulses were intentionally distributed at several onset times to test the contention that a drift bias signal may increase as the trial unfolds (Hanks et al., 2011). An additional analysis was conducted to ensure that the failure to detect a cue effect in the CPP pulse response could not be explained by a dynamic bias that only exerts a significant force at later stages of the trial, but once more, there was no evidence of an expectational effect on the CPP response to the pulses.

In addition to experimentally controlled expectations based on cues or stimulus frequencies, expectations can also be shaped by endogenous sources, like choice history biases. This research suggests that expectations based on the history of decisions made on previous trials (e.g. Urai et al., 2019) or the state of the decision process within a trial (e.g. Talluri et al., 2018) can influence the rate of evidence accumulation. In the present study, an analysis of the CPP slope immediately following a pulse indicated that the slope was increasingly positive as pulse onset time increased. One interpretation of this is that positive pulses have a greater effect later in the trial, but gap and reverse pulses have a weakening influence, independent of cue condition. This might suggest that a dynamic drift bias does influence decision formation, but that it reflects the weight of accumulated evidence rather than the prior. This kind of confirmation bias effect was previously reported by Talluri et al. (2018), who showed that evidence consistent with an initial choice about a stimulus is more heavily weighted in subsequent appraisals of the stimulus, possibly through the influence of selective attention. This is consistent with the present study, where, at later stages in the trial, the participant had been exposed to more evidence in the correct direction. A drift bias reflecting this would be expected to enhance the impact of positive pulses and diminish the impact of gap and reverse pulses. In a recent study, Peixoto et al. (2021) reported exactly the opposite effect in monkeys performing random dot motion discrimination. By decoding a decision variable correlate from real time neural activity, they could systematically trigger a pulse of evidence as the decoded decision variable hit imposed virtual limits. They found that the pulses led to larger changes in the decoded decision variable at lower limits than when there had been significant evidence accumulation. They attributed this to the decision variable being closer to the bound at higher virtual thresholds. Indeed, this effect has previously been observed in a number of monkey decision-making studies (Huk & Shadlen, 2005; Kiani et al., 2008; Mazurek et al., 2003). It was suspected that this issue also affected several analyses of the influence of pulse onset in this study (Figure 6.6.B). However, it is not clear that the CPP's build-up necessarily stops at the time that motor thresholds are passed, and previous work has demonstrated that the CPP can, in certain

circumstances, continue to accumulate evidence after choice commitment (Murphy et al., 2015). Additionally, Peixoto et al. focussed on the amplitude of the decoded decision variable, which would be more sensitive to bound proximity than slope measurements, and this in turn could plausibly mask the kind of confirmation bias effect reported by Talluri and colleagues.

The Complication of Urgency

One factor that could complicate the interpretation of some of these effects is the presence of a dynamic urgency signal. Murphy et al. (2016) describe three hallmarks of urgency: few missed deadlines, strongly negative conditional accuracy functions, and chance performance around the time of the response deadline. The average valid (not too early or late) response rate in this dataset was 97.5%; as shown in Figure 6.8.A, accuracy in the neutral cue condition had a strong negative trend across reaction times; and the mean accuracy for reaction times in the final 200 ms before the response deadline, which constitute only 3% of trials, was 49.91%. In addition, the build-up of motor preparation signals is apparent in advance of evidence onset, which is regarded as a key indicator of dynamic urgency (Churchland et al., 2008; Hanks et al., 2014). Indeed, this was one of the observations that led Kelly et al. (2021) to include an urgency component in their neurally-informed model, which indicated that the influence of a predictive cue was best explained through the inclusion of a drift bias parameter.

Although the CPP amplitude at response did not decrease as reaction time increased in the present study, it is useful to compare these results to those reported by Kelly et al. (2021). In their 'Deadline' condition, which had strong evidence and a short response deadline, they also found increasing CPP amplitudes at response as reaction times increased. In contrast, in their 'LoCoh' condition, which is most comparable to this study, the amplitude of the response-locked CPP appeared to be unaffected by the cue conditions. This difference between the studies may be attributable to the different stimuli or the number of interleaved regimes used by Kelly and colleagues. For example, it is possible that prior probability is a more persistent influence when participants come to expect more volatility in difficulty and response deadline across blocks, compared to the constant task parameters in this study. In fact, Hanks et al. (2011) specifically suggested that the dynamic bias signal should be implemented when there is uncertainty about the reliability of the sensory evidence. Practice could also be a relevant factor in this context, Kelly et al.'s participants completed two testing sessions, but the participants in this study completed between four and six sessions. It is possible that greater exposure to the task allowed participants to more accurately estimate the average reliability of the sensory evidence, which could have several consequences, including undermining the value of a dynamic bias signal.

Importantly, each of the response-locked CPP patterns reported by Kelly et al. emerged from the complex interplay of urgency, starting point biases, and drift biases that were only rendered interpretable with modelling (Figure 6.2.B). While it is possible that the result does simply reflect higher bounds at later stages of the trial, and some have proposed accounts where the decision thresholds can expand across the trial (Deneve, 2012),

in the context of the literature this is not the most likely explanation. It seems clear that drawing firm conclusions about the role of drift bias in the present results will only be made possible by parsing some of this ambiguity with computational modelling.

Summary

In conclusion, there is strong support for a starting point bias in both the behavioural and neurophysiological analyses, but there are also several indications that the starting point may not be the only mechanism representing the influence of expectation in the decision process. The reaction time distributions, sensory-level modulation, and pre-evidence motor preparation build-up are all indications that additional mechanisms may be required to account for the role of priors. The mapping between model parameters and the dynamics of neural signals is not always reliable and a number of studies have illustrated the perils of inferring model adjustments directly from behavioural or neural data (Purcell et al., 2017; Voss et al., 2004). However, there is precedent in the approach taken by Kelly et al. (2021), demonstrating that complex adjustments, which could not be understood with direct interrogation of the empirical data alone, can be explored and interpreted by carefully tailoring a model to the neurophysiological data. Therefore, this chapter represents the first phase of analysis for this dataset. It is hoped that some of the questions that inspired the project, and have eluded explanation in the present analyses, can be resolved with neurally-informed modelling in the future.

References

Aitchison, L., & Lengyel, M. (2017). With or without you: predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, *46*, 219–227.

Albright, T. D. (2012). On the perception of probable things: neural substrates of associative memory, imagery, and perception. *Neuron*, *74*, 227–245.

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex*, *80*, 125–140.

Bar, M. (2004). Visual objects in context. *Nature Reviews. Neuroscience*, *5*, 617–629.

Basso, M. A., & Wurtz, R. H. (1997). Modulation of neuronal activity by target uncertainty. *Nature*, *389*, 66–69.

Basso, M. A., & Wurtz, R. H. (1998). Modulation of neuronal activity in superior colliculus by changes in target probability. *The Journal of Neuroscience*, *18*, 7519–7534.

Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J., & Ungerleider, L. G. (2016). Encoding of stimulus probability in macaque inferior temporal cortex. *Current Biology*, *26*, 2280–2290.

Bell, A. H., Summerfield, C., Morin, E. L., Malecek, N. J., & Ungerleider, L. G. (2017). Reply to vinken and vogels. *Current Biology*, *27*, R1212–R1213.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*, 700–765.

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*, 693–702.

Deneve, S. (2012). Making decisions with unknown sensory reliability. *Frontiers in Neuroscience*, *6*, 75.

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*, 764–779.

de Lange, F. P., Rahnev, D. A., Donner, T. H., & Lau, H. (2013). Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *The Journal of Neuroscience*, *33*, 1400–1410.

Devine, C. A., Gaffney, C., Loughnane, G. M., Kelly, S. P., & O'Connell, R. G. (2019). The role of premature evidence accumulation in making difficult perceptual decisions under temporal uncertainty. ELife, 8.

Diederich, A., & Busemeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception & Psychophysics*, *68*, 194–207.

Donner, T. H., Siegel, M., Fries, P., & Engel, A. K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Current Biology*, *19*, 1581–1585.

Dorris, M. C., & Munoz, D. P. (1998). Saccadic probability influences motor preparation signals and time to saccadic initiation. *The Journal of Neuroscience*, *18*, 7015–7026.

Dunovan, K. E., Tremel, J. J., & Wheeler, M. E. (2014). Prior probability and feature predictability interactively bias perceptual decisions. *Neuropsychologia*, *61*, 210–221.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *The Journal of Neuroscience*, *30*, 16601–16608.

Feuerriegel, D., Blom, T., & Hogendoorn, H. (2021b). Predictive activation of sensory representations as a source of evidence in perceptual decision-making. Cortex, 136, 140–146.

Feuerriegel, D., Churches, O., Coussens, S., & Keage, H. A. D. (2018). Evidence for spatiotemporally distinct effects of image repetition and perceptual expectations as measured by event-related potentials. *Neuroimage*, *169*, 94–105.

Feuerriegel, D., Vogels, R., & Kovács, G. (2021a). Evaluating the evidence for expectation suppression in the visual system. Neuroscience and Biobehavioral Reviews, 126, 368–381.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. *Annual Review of Psychology*, *67*, 641–666.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*, 127–138.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

Gold, J. I., & Stocker, A. A. (2017). Visual Decision-Making in an Uncertain and Dynamic World. *Annual Review of Vision Science*, *3*, 227–250.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*, 14–23.

Grotheer, M., & Kovács, G. (2016). Can predictive coding explain repetition suppression? *Cortex*, *80*, 113–124.

Hanks, T., Kiani, R., & Shadlen, M. N. (2014). A neural mechanism of speed-accuracy tradeoff in macaque area LIP. ELife, 3.

Hanks, T. D., Ditterich, J., & Shadlen, M. N. (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nature Neuroscience*, *9*, 682–689.

Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *The Journal of Neuroscience*, *31*, 6339–6352.

Heeger, D. J. (2017). Theory of cortical function. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 1773–1782.

Heilbron, M., & Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, *389*, 54–73.

Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *The Journal of Neuroscience*, *25*, 10420–10436.

Kaliukhovich, D. A., & Vogels, R. (2011). Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cerebral Cortex*, *21*, 1547–1558.

Kelly, S. P., Corbett, E. A., & O'Connell, R. G. (2021). Neurocomputational mechanisms of prior-informed perceptual decision-making in humans. *Nature Human Behaviour*, *5*, 467–481.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *The Journal of Neuroscience*, *33*, 19434–19441.

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *The Journal of Neuroscience*, *28*, 3017–3029.

Kim, Y. J., Grabowecky, M., Paller, K. A., Muthu, K., & Suzuki, S. (2007). Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nature Neuroscience*, *10*, 117–125.

Kok, P., Brouwer, G. J., van Gerven, M. A. J., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *The Journal of Neuroscience*, *33*, 16275–16284.

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of Cognitive Neuroscience*, *26*, 1546–1554.

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, *75*, 265–270.

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 10473–10478.

Kovács, G., Kaiser, D., Kaliukhovich, D. A., Vidnyánszky, Z., & Vogels, R. (2013). Repetition probability does not affect fMRI repetition suppression for objects. *The Journal of Neuroscience*, *33*, 9805–9812.

Kubanek, J., Snyder, L. H., Brunton, B. W., Brody, C. D., & Schalk, G. (2013). A low-frequency oscillatory neural signal in humans encodes a developing decision variable. *Neuroimage*, *83*, 795–808.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmertest package: tests in linear mixed effects models. Journal of Statistical Software, 82.

Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cerebral Cortex*, *22*, 567–576.

Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgement and Decision Making*, *6*, 651–687.

Lochmann, T., & Deneve, S. (2011). Neural processing as causal inference. *Current Opinion in Neurobiology*, *21*, 774–781.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494–1502.

Martinez-Rodriguez, L. A., Corbett, E. A., O'Connell, R. G., & Kelly, S. P. (2020). Neurally-Informed Modelling of Static and Dynamic Decision Biases. *Research Gate*.

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex*, *13*, 1257–1269.

McGovern, D. P., Hayes, A., Kelly, S. P., & O'Connell, R. G. (2018). Reconciling age-related changes in behavioural and neural indices of human perceptual decision-making. *Nature Human Behaviour*, *2*, 955–966.

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: a diffusion model analysis of prior probability and potential payoff. The Journal of Neuroscience, 32, 2335–2343.

Mulder, M. J., van Maanen, L., & Forstmann, B. U. (2014). Perceptual decision neurosciences - a model-based review. Neuroscience, 277, 872–884.

Murphy, P. R., Boonstra, E., & Nieuwenhuis, S. (2016). Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nature Communications*, *7*, 13526.

Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. ELife, 4.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*, 1729–1735.

O'Connell, R. G., & Kelly, S. P. (2021). Neurophysiology of Human Perceptual Decision-Making. *Annual Review of Neuroscience*, *44*, 495–516.

Peixoto, D., Verhein, J. R., Kiani, R., Kao, J. C., Nuyujukian, P., Chandrasekaran, C., … Newsome, W. T. (2021). Decoding and perturbing decision states in real time. *Nature*, *591*, 604–609.

Platt, M. L. (2002). Neural correlates of decisions. Current Opinion in Neurobiology, 12, 141–148.

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. Nature, 400, 233–238.

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, *16*, 1170–1178.

Purcell, B. A., & Palmeri, T. J. (2017). Relating accumulator model parameters and neural dynamics. *Journal of Mathematical Psychology*, *76*, 156–171.

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99.

Rao, V., DeAngelis, G. C., & Snyder, L. H. (2012). Neural correlates of prior expectations of motion in the lateral intraparietal and middle temporal areas. *The Journal of Neuroscience*, *32*, 10063–10074.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367.

Richter, D., & de Lange, F. P. (2019). Statistical learning attenuates visual activity only for attended stimuli. ELife, 8.

Rungratsameetaweemana, N., & Serences, J. T. (2019). Dissociating the impact of attention and expectation on early sensory processing. *Current Opinion in Psychology*, *29*, 181–186.

Schlack, A., & Albright, T. D. (2007). Remembering visual motion: neural correlates of associative plasticity and motion recall in cortical area MT. *Neuron*, *53*, 881–890.

Shulman, G. L., Ollinger, J. M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Petersen, S. E., & Corbetta, M. (1999). Areas involved in encoding and applying directional expectations to moving objects. *The Journal of Neuroscience*, *19*, 9480–9496.

Siegel, M., Engel, A. K., & Donner, T. H. (2011). Cortical network dynamics of perceptual decision-making in the human brain. *Frontiers in Human Neuroscience*, *5*, 21.

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Current Opinion in Neurobiology*, *13*, 144–149.

Steinemann, N. A., O'Connell, R. G., & Kelly, S. P. (2018). Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nature Communications*, *9*, 3627.

Summerfield, C., & Blangero, A. (2017). Perceptual Decision-Making. In Decision Neuroscience (pp. 149–162). Elsevier.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. Nature Reviews. Neuroscience, 15, 745–756.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. Trends in Cognitive Sciences, 13, 403–409.

Summerfield, C., & Egner, T. (2016). Feature-Based Attention and Feature-Based Expectation. Trends in Cognitive Sciences, 20, 401–404.

Summerfield, C., Wyart, V., Johnen, V. M., & de Gardelle, V. (2011). Human scalp electroencephalography reveals that repetiti suppression varies with expectation. Frontiers in Human Neuroscience, 5, 67.

Talluri, B. C., Braun, A., & Donner, T. H. (2021). Decision making: How the past guides the future in frontal cortex. *Current Biology*, *31*, R303–R306.

Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M., & Donner, T. H. (2018). Confirmation Bias through Selective Overweighting of Choice-Consistent Evidence. *Current Biology*, *28*, 3128-3135.e8.

Teufel, C., & Fletcher, P. C. (2020). Forms of prediction in the nervous system. *Nature Reviews. Neuroscience*, *21*, 231–242.

Trapp, S., Lepsien, J., Kotz, S. A., & Bar, M. (2016). Prior probability modulates anticipatory activity in category-specific areas. *Cognitive, Affective & Behavioral Neuroscience*, *16*, 135–144.

Tzagarakis, C., Ince, N. F., Leuthold, A. C., & Pellizzer, G. (2010). Beta-band activity during motor planning reflects response uncertainty. *The Journal of Neuroscience*, *30*, 11270–11277.

Utzerath, C., St John-Saaltink, E., Buitelaar, J., & de Lange, F. P. (2017). Repetition suppression to objects is modulated by stimulus-specific expectations. *Scientific Reports*, *7*, 8781.

Vinken, K., & Vogels, R. (2017). Adaptation can explain evidence for encoding of probabilistic information in macaque inferior temporal cortex. *Current Biology*, *27*, R1210–R1212.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: an empirical validation. *Memory & Cognition*, *32*, 1206–1220.

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*, 242–268.

Wilming, N., Murphy, P. R., Meyniel, F., & Donner, T. H. (2020). Large-scale dynamics of perceptual decision information across human cortex. *Nature Communications*, *11*, 5109.

Appendix 6.1 - Supplementary Figures

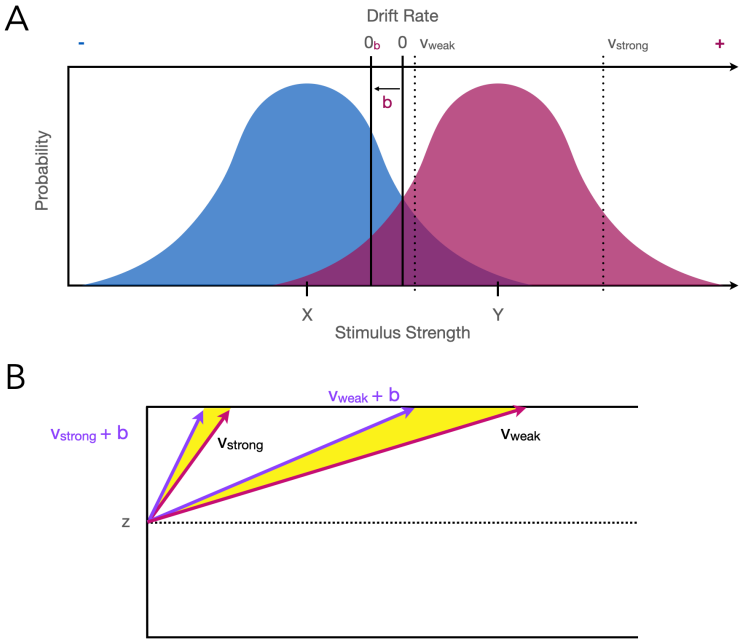


Figure 6.S1. *Drift rate, drift criterion, and drift bias. In this example, the blue stimulus has lower intensities, centred on mean X, and the pink stimulus has higher intensities, centred on mean Y. The blue stimulus has been arbitrarily mapped to the lower bound, and the pink stimulus to the upper bound, in the Drift Diffusion Model. A) The drift criterion (0) segregates the sensory input that produces negative or positive drift rates, favouring the blue or pink stimulus respectively. When there is no expectation that one stimulus is more likely than another, the drift criterion (0) should be set so that there will be a zero drift rate at the stimulus strength where the distributions associated with each alternative intersect (i.e. where there is no informative sensory evidence). A drift bias (b) favouring the pink stimulus offsets this criterion (0b), such that the same stimulus strength produces a more favourable drift rate for the pink stimulus and stronger evidence in favour of the blue stimulus is needed to achieve the same drift rate. B) The two examples of drift bias have equal magnitude (the angle between each and the corresponding unbiased trajectory is equal). However, as the decision variable radiates out from the starting point (z), the impact of the bias on response time is much larger for the weak stimulus than the strong stimulus.*

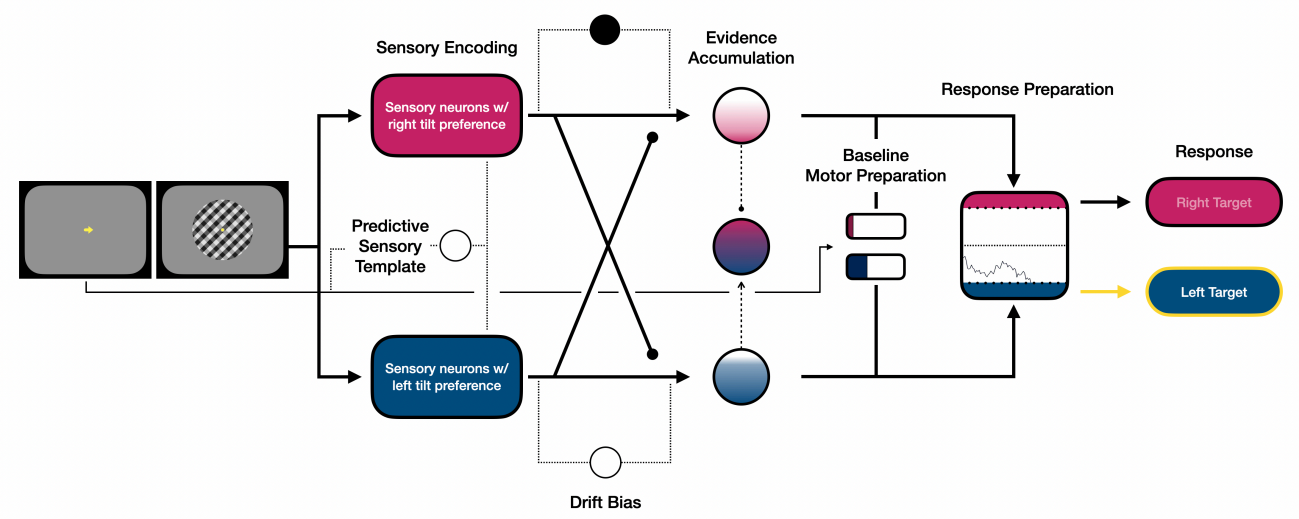


Figure 6.S2. *A tentative schematic of the decision process adapted from Kelly et al. (2021). Sensory evidence from a cued tilted grating stimulus is encoded by populations of orientation-selective neurons in early visual cortex. If preparatory sensory activity contributes to the decision in the form of a 'predictive sensory template', this process may begin before evidence onset. The relative sensory evidence favouring each alternative is continuously fed into separate accumulators at an effector-independent, abstract decision level. The graded colours represent the noisiness of the encoded sensory evidence. The CPP is thought to originate from the combined activity of these accumulators, representing the relative cumulative evidence (central circle). If a drift bias contributes to the decision, evidence for the expected alternative may be enhanced at the sensory level or the drift criterion may shift to bias the interpretation of sensory input at the evidence accumulation stage. The drift bias has opposite effects for the two alternatives. Prior expectations from the cue engender preparatory motor activity favouring the more probable response, this baseline activity is combined with the integrated evidence at the response preparation stage, where the alternatives compete to reach a decision threshold. The relative preparation for each response can be modelled as a Drift Diffusion process. Here, it is clear that the baseline differences in motor preparation manifest as a starting point bias favouring a left response. An equal increase in the baseline motor preparation for both alternatives would reduce the quantity of evidence needed by either alternative to reach the threshold (represented by the white space in baseline motor preparation), so this would be equivalent to reducing the bounds in the Drift Diffusion Model. When enough evidence is accumulated to match the deficit between the starting point and the decision threshold, a response is initiated (yellow surround).*

Appendix 6.2 - Using The Gap Pulse As A Control Condition

To assuage concerns about the no-pulse slicing approach adopted in the analyses reported in this chapter, the analyses were repeated, excluding the no pulse trials and using the gap pulse as the control condition instead. As shown below, this did not meaningfully change any of the results; any differences are highlight in red. The same paragraphs are reproduced from the results section with the stats updated, so the analyses can be easily searched.

The Effect of Cues and Pulses on Reaction Time and Accuracy

The same analysis was carried out to determine the impact of the pulses on reaction time for correct and error response trials. There was a main effect of accuracy ($F(1,11) = 22.678$, $p < 0.001$) and pulse-type ($F(1.114,12.254) = 20.041$, $p < 0.001$), and a significant interaction between accuracy and pulse-type ($F(2,22) = 14.315$, $p < 0.001$). Pairwise comparisons showed that reaction times were faster on correct response trials ($\Delta = -43\pm20$ ms, $p < 0.001$). Further comparisons indicated that positive pulses led to significantly faster responses than gap pulse trials ($\Delta = -37\pm15$ ms, $p = 0.02$). There was no significant difference between reaction times on gap and reverse pulse trials ($\Delta = 0\pm7$ ms, $p = 0.968$). Correct and error responses were analysed separately to characterise the interaction. The was a main effect of pulse-type for correct responses ($F(2,22) = 36.278$, $p < 0.001$) was further analysed with pairwise comparisons. Reaction times were significantly faster following positive pulses ($\Delta = -63\pm24$ ms, $p < 0.001$) and significantly slower following reverse pulses ($\Delta = 15\pm14$ ms, $p = 0.035$), compared to the gap pulse condition. There was no main effect of pulse for error responses ($F(1.208,13.289) = 1.093$, $p = 0.329$).

For correct response trials, there was a main effect of pulse-type ($F(2,36100.1) = 158.743$, $p < 0.001$) and cue condition ($F(2,36101.388) = 133.148$, $p < 0.001$), but there was no interaction ($F(4,36100.077) = 0.937$, $p = 0.441$). Pairwise comparisons showed that, compared to neutral cues, valid cues significantly reduced reaction times ($\Delta = -29\pm8$ ms, $p < 0.001$) and invalid cues significantly increased reaction times ($\Delta = 41\pm11$ ms, $p < 0.001$). In addition, positive pulses reduced reaction times ($\Delta = -78\pm9$ ms, $p < 0.001$), while reverse pulses ($\Delta = 19\pm10$ ms, $p< 0.001$) slowed reaction times, compared to the gap pulse condition.

For error responses trials, there was a main effect of cue type ($F(2,10832.004) = 74.914$, $p < 0.001$), but there was no main effect of pulse type ($F(2,10827.515) = 2.053$, $p = 0.128$), and no interaction ($F(4,10826.254) = 0.743$, $p = 0.563$). Pairwise comparisons showed that valid cues significantly increased reaction times ($\Delta = 40\pm14$ ms, $p < 0.001$) and invalid cues significantly reduced reaction times ($\Delta = -39\pm15$ ms, $p < 0.001$).

The analyses of the effect of the pulse across pulse onsets were not reproduced because the conclusion of those analyses was that they were ill-equipped to really shed light on the hypothesis being tested and that the question of a dynamic drift bias will have to be addressed in future modelling work.

A repeated measures ANOVA was conducted on the accuracy data to assess the influence of the cues and pulses, using cue condition and pulse-type as factors. There was a main effect of cue condition ($F(1.128,12.407) = 46.392$, $p < 0.001$), and pulse-type ($F(1.068,11.744) = 112.701$, $p < 0.001$), and a significant interaction ($F(2.016,22.172) = 14.362$, $p < 0.001$). Pairwise comparisons showed that accuracy was higher on valid cue trials ($\Delta = 15.7\pm5\%$, $p < 0.001$) and lower on invalid cue trials ($\Delta = -16\pm5.9\%$, $p < 0.001$), compared to the neutral cue condition. Positive pulses increased accuracy ($\Delta = 11.6\pm2.1\%$, $p < 0.001$), while reverse pulses reduced accuracy ($\Delta = -3.4\pm1.2\%$, $p < 0.001$), compared to the gap pulse condition. The interaction was investigated with separate ANOVAs for each of the pulses baselined by the gap pulse condition. There was a main effect of cue condition in the baselined positive pulse data ($F(1.226,13.481) = 23.731$, $p < 0.001$). The effect of the positive pulse was significantly reduced in the valid cue condition ($\Delta = -6.2\pm2.3\%$, $p < 0.001$) and significantly greater in the invalid cue condition than the neutral cue condition ($\Delta = 4.1\pm2.9\%$, $p = 0.01$). There was no main effect of cue condition in the baselined reverse pulse data ($F(2,22) = 0.881$, $p = 0.429$).

* This analysis was conducted simply to illustrate the mutual influence of the cues and pulses on reaction times and this comparison was very nearly significant in the original analysis ($p = 0.056$). This change has no broader implications for the interpretation of the results.

The Effect of Priors on the Encoding of Sensory Evidence

A mixed effects analysis of the pulse-locked marginal SSVEP in the window 0:250 ms was conducted, using SSVEP target frequency, pulse-type, cue condition, and an interaction between pulse-type and cue condition as predictors. The marginal SSVEP was re-baselined in the window -250:-50 ms to slate the pulse effect. There was a significant main effect of pulse type ($F(2,32022.581) = 314.414$, $p < 0.001$), but no main effect of cue condition ($F(2,32023.618) = 0.078$, $p = 0.925$) and no interaction between pulse type and cue condition ($F(4,32021.798) = 0.73$, $p = 0.572$). There was also no main effect of SSVEP frequency ($F(1,32029.956) = 1.812$, $p = 0.178$). Pairwise comparisons showed that, compared to the gap pulse condition, positive pulses significantly increased the marginal SSVEP ($p < 0.001$), and reverse pulses significantly reduced the marginal SSVEP ($p < 0.001$).

There was a main effect of pulse type ($F(2,32019.8) = 188.98$, $p < 0.001$) and pulse onset ($F(1,32022.276) = 23.248$, $p < 0.001$), but no main effect of cue condition ($F(2,32020.281) = 0.048$, $p = 0.953$). There was a significant interaction between pulse type and pulse onset time ($F(2,32019.53) = 70.054$, $p < 0.001$), but no cue-pulse interaction ($F(4,32019.365) = 0.688$, $p = 0.6$). The interaction was characterised with separate mixed effects analyses for each pulse type. The effect of a positive pulse significantly decreased as onset time

increased ($F(1,10564.922) = 44.968$, $\beta = -2.73$, $p < 0.001$). The effect of a gap pulse significantly increased across onset times ($F(1,10722.382) = 63.531$, $\beta = 3.085$, $p < 0.001$). The effect of a reverse pulse also significantly increased across onset times ($F(1,10718.772) = 55.392$, $\beta = 2.936$, $p < 0.001$).

* It is unsurprising that when the no pulse condition was removed from the analysis the main effect of pulse onset became significant, since all pulse-types individually changed significantly across pulse onset in the original analysis. This change doesn't not have any important implications for the original analysis.

The Effect of Priors on the Accumulation of Evidence

In the analysis of the CPP amplitude, there was a main effect of pulse-type ($F(2,34120.873) = 25.433$, $p < 0.001$) and a significant interaction between pulse-type and accuracy ($F(2,34123.58) = 15.955$, $p < 0.001$). There was no main effect of cue condition ($F(2,34124.969) = 1.958$, $p = 0.141$) or accuracy ($F(1,34053.41) = 1.959$, $p = 0.162$), and no interaction between cue condition and accuracy ($F(2,34072.968) = 0.288$, $p = 0.75$). Crucially, there was no interaction between pulse-type and cue condition ($F(4,34118.765) = 0.843$, $p = 0.498$) and no three-way interaction ($F(4,34119.565) = 1.371$, $p = 0.241$). Pairwise comparisons showed that, compared to the gap pulse condition, CPP amplitude was significantly greater following a positive pulse ($p = 0.001$), but there was no significant difference in CPP amplitude on reverse and gap pulse trials ($p = 0.736$). The relationship between pulse-type and accuracy will be investigated Chapter 4, so the interaction was not explored here.

In the analysis of pulse-locked CPP slope, there was a main effect of pulse-type ($F(2,34124.501) = 46.624$, $p < 0.001$) and accuracy ($F(1,32826.863) = 5.612$, $p = 0.018$), there was also a significant interaction between pulse-type and accuracy ($F(2,34125.752) = 34.481$, $p < 0.001$). There was no main effect of cue ($F(2,34118.656) = 0.035$, $p = 0.965$) and no interaction between cue and accuracy ($F(2,33306.33) = 1.083$, $p = 0.338$). Importantly, there was also no interaction between cue and pulse-type ($F(4,34120.535) = 0.745$, $p = 0.561$), and no three-way interaction ($F(4,34122.241) = 1.507$, $p = 0.197$). Pairwise comparisons showed that, compared to the gap pulse condition, the CPP slope was significantly steeper following a positive pulse ($p < 0.001$). There was no significant difference in CPP slope in reverse pulse ($p = 0.443$) compared to gap pulse trials. Pairwise comparisons also showed that the pulse-locked CPP slope was significantly steeper on error response trials ($p = 0.018$). Once more, the relationship between pulse-type and accuracy will be investigated in the next chapter, so the interaction was not explored here.

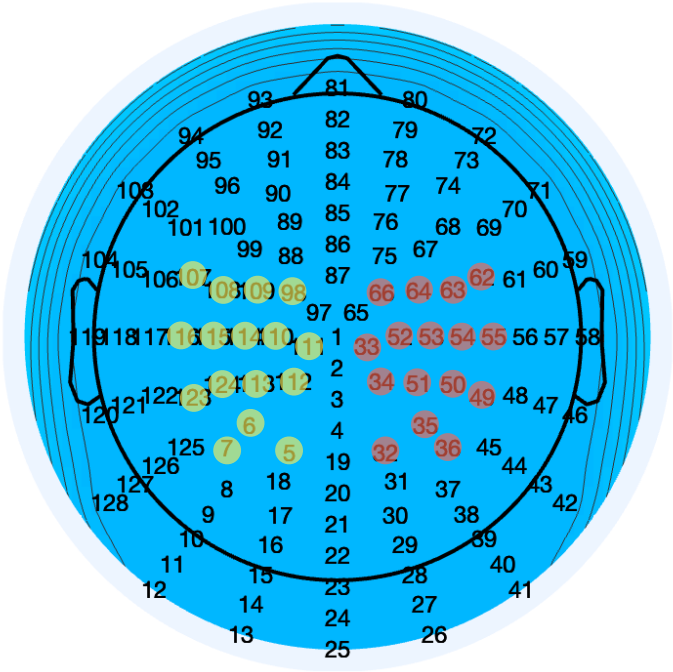
The only significant result of the CPP amplitude analysis was a main effect of pulse-type ($F(2,25935.687) = 17.265$, $p < 0.001$). There was no significant main effect of cue condition ($F(2,25938.896) = 0.707$, $p = 0.493$) or pulse onset ($F(1,25936.093) = 0.049$, $p = 0.824$). There was also no significant interaction between cue condition and pulse-type ($F(4,25935.084) = 0.659$, $p = 0.621$) or between cue condition and pulse onset ($F(2,25935.088) = 0.022$, $p = 0.978$) or between pulse-type and pulse onset ($F(2,25934.773) = 0.863$, $p = 0.422$), and no three-way interaction ($F(4,25934.84) = 0.292$, $p = 0.883$). Pairwise comparisons showed that,

compared other gap pulse condition, CPP amplitude was significantly greater following a positive pulse ($p < 0.001$). There was no significant difference between reverse pulse and gap pulse trials ($p = 0.088$).

In the analysis of CPP slope, there was a main effect of pulse-type ($F(2,25939.855) = 48.796$, $p < 0.001$) and pulse onset ($F(1,25941.48) = 9.014$, $p = 0.003$). There was no main effect of cue condition ($F(2,25935.046) = 0.138$, $p = 0.871$), and no interaction between pulse-type and cue condition ($F(4,25937.627) = 1.454$, $p = 0.213$), cue condition and pulse onset ($F(2,25937.901) = 0.066$, $p = 0.936$), or pulse-type and pulse onset ($F(2,25937.11) = 0.195$, $p = 0.823$). There was also no three-way interaction ($F(4,25936.983) = 0.911$, $p = 0.456$). Pairwise comparisons showed that, compared to the gap pulse condition, CPP slope was significantly greater following a positive pulse ($p < 0.001$). There was no significant difference between gap and reverse pulses ($p = 0.291$). The main effect of onset time was interpreted by repeating the analysis, but removing all interaction terms. CPP slope increased as onset time increased ($F(1,25956.952) = 17.402$, $\beta = 0.106$, $p < 0.001$).

* The first changed result is of no importance in that analysis. The only result of importance would have been the interaction between pulse-type and cue condition or the three-way interaction. The second changed result is also not relevant to the purpose of that analysis. The pulse effect on CPP slope has been established in a previous analysis. Instead, the intention here was to test for a cue-pulse interaction and that result remains unchanged.

Appendix 6.3 - Electrode Selection



The pool for selecting Mu-Beta electrodes is highlighted above. The individual electrode selections for each participant are listed below.

Left Hemisphere	Right Hemisphere
113, 124	50, 49
113,124	50, 51, 53, 54
114, 115	54, 55
110, 112	53, 54, 55
115, 114	54, 55
115, 116	54, 53
114, 115, 124	53, 54, 63, 64
109, 110	49, 50
123, 124, 6	51, 50
123, 116, 115	53, 64, 63
114, 115	54, 55
112	49, 54

7.

General Discussion

Bertrand Russell famously considered an intuitive response to the question of whether all knowledge can be doubted: that perceptual experience provides an objective and uncontroversial kind of knowledge. He described surveying the objects in his study and noting that it was unlikely an impartial observer would doubt the contention that his table was brown. However, on closer inspection, some parts of the table were in shade and other parts were reflecting light. If the observer attributed a specific hue to any part of the table, flicking a light switch would be a damning rebuttal. In fact, any pattern of retinal stimulation can be caused by many stimulus configurations, and a single stimulus, like a table, can provoke many distinct patterns of retinal activity. The time of day, the tint of the lighting, the polish on the table, the observer's spectacles, and their position in the room are just some of the causes that conspire to give rise to the particular percept of colour. It is for this reason that Helmholtz described perception as a process of unconscious inference (Wade, 2021). The sensory apparatus does not have access to the hidden causes generating sequences of retinal excitation or reverberations in the inner ear, instead the brain must assemble coherent perceptual experience from noisy sensory data by reconstructing its likely causes.

Selecting a perceptual hypothesis appears to be an impossible problem when more than one permutation of these causes can produce identical sensory data, but entertaining simultaneous conflicting percepts or deliberating ad infinitum would render any animal helpless in a hostile environment. Instead, this uncertainty must be constrained by integrating the noisy sensory input with previously acquired knowledge about the world (Kording, 2014; Summerfield & de Lange, 2014). If photons independently jostled all of the

photoreceptors in the human retina between one of two states - on or off - the possible variations of retinal input would outnumber atoms in the visible universe by $10^{10,000,000}$ (Gold & Stocker, 2017). Fortunately, visual information is not independently distributed in natural scenes, there are general statistical regularities found in the spatiotemporal distribution of low-level visual features like orientation, spatial frequency, and contrast (Geisler, 2008). As described by Horace Barlow's (1961) Efficient Coding Hypothesis, evolutionary forces reined in the apparent computational intractability of resolving the ambiguity in the sensory stream by leveraging these statistical regularities in the brain's processing architecture (Barlow, 1972; Ganguli & Simoncelli, 2014), optimising the use of limited neural resources to encode a robust and adaptive perceptual experience (Geisler, 2008; Hermundstad et al., 2014; Simoncelli, 2003; Simoncelli & Olshausen, 2001; Teufel & Fletcher, 2020).

For example, cardinal orientations are disproportionately represented in natural images and this statistical inequality is mirrored in the anatomy of primary visual cortex (Girshick et al., 2011); V1 responds more vigorously to cardinal orientations, more cells in V1 are tuned to cardinal orientations than oblique ones, and the tuning functions of these cells are narrower than for those with oblique preferences (Furmanski & Engel, 2000; Li et al., 2003). The implicit representation of this prior information in neurophysiology is reflected in behaviour; psychophysical thresholds for orientation discrimination, contrast sensitivity, and motion detection, among other tasks, are lower for stimuli aligned to cardinal axes (Girshick et al., 2011; Mitchell et al., 1967; Stein & Peelen, 2015). While it is assumed that these structural adaptations are primed by evolution, as in all aspects of perceptual processing, experience is critical (Daw, 2009; Wiesel & Hubel, 1963). Even perceptual biases based on regularities as reliable as the sun's propensity to shine from above can be altered with experience (Adams et al., 2004). Berkes et al. (2011) suggested that with greater exposure to natural scenes endogenous activity in visual cortex should increasingly represent this internal model of the sensory environment. They found that over the course of ferrets' development, the distributions of spontaneous V1 activity in darkness and V1 responses to natural scenes became increasingly similar, but responses to artificial stimuli remained entirely distinct.

The influence of priors is not just confined to evolutionary and ontogenetic timescales. Prior experiences are a critical factor in moment-by-moment perceptual decisions (Gold & Stocker, 2017; Series & Seitz, 2013), as statistical regularities also infuse daily visual experience: objects tend to be static or move relatively slowly along predictable trajectories, object identities are typically stable over time, and particular objects are more probable in certain contexts than others (Bar, 2004; Stocker & Simoncelli, 2006; Weiss et al., 2002). For example, when you turn on the light in a kitchen, you are more likely to find a table than a leopard. Such contextual probabilities can constrain perceptual uncertainty by reducing the pool of candidate interpretations. Indeed, object recognition is faster and more accurate when objects are placed in familiar contexts (Bar, 2004; Bar & Ullman, 1996; Palmer, 1975), but recognition is slower and errors are more common when objects' position, size, or identity conflicts with the background context (Biederman et al., 1982), or when the contextual information is scrambled (Biederman, 1972). Humans are well-tuned to extract prior information that may facilitate these kinds of perceptual inferences based on stimulus frequencies, associative learning, transitional probabilities, and contextual processing (Bar, 2004; Summerfield & Egner, 2009; Turk-Browne et al., 2010; Zhao

et al., 2013). By habituating infants to stimulus pairings or visual events and then violating the associated expectations, researchers have shown that the ability to track these perceptual dependencies is evident in the behavioural responses of children as young as 2.5 months (Baillargeon, 2004; Fiser & Aslin, 2002). In adults, these regularities modulate the neurophysiological processing of stimulus sequences (de Lange et al., 2018; de Lange & Summerfield, 2014; Walsh et al., 2020), even when the participants have no awareness of the pattern (Turk-Browne et al., 2009, 2010) or the pattern is embedded in complex naturalistic stimuli (Maniscalco et al., 2018).

It appears that at every level of processing, prior experience contributes to the construction of the perceptual world. In the last two decades, there has been renewed interest in the mechanisms that implement these contributions and the implications for the broader understanding of the processing stream. Some of this interest has been spurred by contentious theoretical accounts claiming that prediction is the canonical and unifying neural function (Friston, 2009, 2010; Walsh et al., 2020), but the advances inspired by this debate also contribute to the expansion of the traditional framework to reflect an increasingly sophisticated role for expectation in its various guises (Bubic et al., 2010; de Lange et al., 2018; Gilbert & Li, 2013; Lochmann & Deneve, 2011; Press et al., 2020; Pouget et al., 2013; Rust & Palmer, 2020; Teufel & Nanay, 2017; Teufel & Fletcher, 2020). This thesis sought to contribute to this growing literature, by examining the influence of prior experience in the early processing of sensory information, the evaluation of that sensory evidence in perceptual decision-making, and the outcome of these processes in perceptual experience. The following sections discuss the results of these investigations, their limitations, and potential directions for future research. The thesis was composed of two halves; the first addressing the contentions of a popular alternative framework that emphasises the role of expectation in sensory processing, and the second examining the mechanisms responsible for expectation effects in perceptual decisions. The General Discussion will reflect this partition, addressing Chapter 2 and 3 in the first section and Chapter 5 and 6 in the second section.

Exploring The Divergence of The Traditional Model of Sensory Processing and Predictive Processing

One of the overarching objectives of neuroscience is to provide an explanation for our subjective experience in terms of its neural substrates. In 1962, David Hubel and Thorsten Wiesel published a landmark paper representing a significant step towards that goal. The paper offered the first description of the role of single neurons in cortical sensory processing. These neurons in primary visual cortex were unresponsive to the single points of light that had been found to excite centre-surround cells in the retina and lateral geniculate nucleus (LGN), instead they exhibited selectivity to a simple sum of these inputs: oriented lines. The recordings of these live cells demonstrated that neural networks could endow specific, complex response properties to individual sensory neurons when they were arranged in a cortical hierarchy (Almasi et al., 2020). Indeed, subsequent research showed that in just the step from the LGN to V1 neural representations of orientation, motion direction, colour, and binocular disparity emerge (Briggs, 2020). Hubel and Wiesel's discovery heralded an era of astonishing advances in the understanding of the anatomical and functional architecture of the visual system (Trenholm & Krishnaswamy, 2020), and laid the foundation for the predominant model of sensory processing (Felleman & Van Essen, 1991; van Essen & Maunsell, 1983). This feedforward model claims that the intricate detail of visual experience is extracted from the hail of noisy sensory stimulation by a processing hierarchy of progressively complex spatiotemporal filters (Bond, 2004; Riesenhuber & Poggio, 1999; van Essen & Gallant, 1994). Each stage along this processing stream assembles more complicated representations from the simpler stimulus properties extracted at the previous stage (Serre, 2015); just as points of light feed representations of oriented lines, oriented lines feed representations of simple shapes and textures. The increasing complexity of neuronal selectivities is paralleled by increasing invariance in the representations, such that neurons at higher-levels of processing are more tolerant of object position, scale, lighting, occlusion, and viewpoint. In this way, a stable representation of the visual scene emerges from the cumulative collation of feature representations across concurrent processing streams as sensory data percolates through the hierarchy, giving rise to an integrated and coherent perceptual experience (Di Carlo et al., 2012; Logothetis & Sheinberg, 1996; Nassi & Calloway, 2009).

This framework has had tremendous success in explaining a vast range of neurophysiological and psychophysical phenomena (Heeger, 2017). So it is perhaps unsurprising, given its monolithic status, that the

proposal of a radical alternative, representing a complete inversion of this processing strategy, was a catalyst in perceptual research. According to this account, termed predictive processing, the statistical regularities that characterise the sensory world are exploited to generate perceptual hypotheses and distribute predictions throughout the processing stream about likely patterns of incoming sensory stimulation (Clark, 2013; Friston, 2005; Hohwy, 2013; Wiese & Metzinger, 2017). When hypotheses are accurate, the predictions are in close accord with the actual sensory input and these feedforward signals are suppressed, but when an umbrella is mistaken for a python, the discrepancy between the sensory predictions and the sensory evidence gives rise to prediction error. Not only does prediction error quantify the adequacy of perceptual hypotheses, more importantly it describes precisely what they got wrong (den Ouden et al., 2012). The prediction error signalling the mismatch between the reptilian expectation and the nylon reality informs a revision of the model generating that prediction to bring its estimate of the sensory causes in line with the sampled data and minimise the prediction error from subsequent predictions. This computational motif of prediction error minimisation is iterated at every stage of the predictive processing hierarchy in the interaction of two neural subpopulations: expectation units and error units. Relevant sensory predictions are carried in cascading feedback to expectation units at each level of analysis to meet sensory signals as they arrive, but only the prediction error arising from these comparisons is propagated forwards to the next stage of processing by error units. In this way, every level of the hierarchy attempts to account for the correspondence of predictions and sensory input at the preceding level and the network gradually converges on an internally consistent account of the sensory environment.

Chapter 2. The Role of Expectation in Binocular Rivalry

This framework makes intriguing predictions about how perception will respond to ambiguous sensory information. Specifically, the predictive processing account of binocular rivalry claims that the contents of perception are determined by a Bayesian weighting of the sensory evidence with prior expectations (Hohwy et al., 2008). This suggests that when the sensory evidence is perfectly matched across perceptual hypotheses, as it is at the onset of binocular rivalry, expectation should be the determining factor in what is perceived at any given moment. This idea was tested in Chapter 2 by exposing participants to a 'prior-display' of oriented gratings with a statistical bias towards one of the orientations that would compete for the participant's perceptual awareness in a subsequent rivalry phase. It was hypothesised that the expectation derived from the regularity in perceptual experience immediately preceding rivalry would prejudice this competition, such that the first percept that the participant reported would be the 'expected' grating and this bias in perceptual dominance would gradually erode as their priors were revised in light of constant and equal sensory evidence presented during rivalry.

However, the results showed the exact opposite pattern. The unexpected grating was more likely to be reported as the first percept and it continued to dominate in the perceptual dynamic during the early phase of

rivalry (~30-40 s). One interpretation of this result is that the more frequent presentation of the 'expected' grating orientation during the prior-display led to an adaptation effect, depressing responses to the expected grating and altering natural rivalry dynamics. A follow-up experiment sought to test whether adaptation could explain this pattern of results by examining whether there was any evidence of biased orientation discrimination responses after exposure to the prior-display. There was no consistent evidence of a tilt aftereffect, but the small sample size of this follow-up experiment may have made detection of adaptation less likely, so this result may be better interpreted as inconclusive. In the future, the tilt aftereffect experiment could be repeated with a larger sample and more testing sessions per participant, but also with the test grating displaced from the central fixation. This may help to prise apart the candidate explanations of a higher-level expectation effect and local adaptation because the tilt aftereffect is specific to a retinotopic area of the visual field (Knapen et al., 2010; Mathot et al., 2013), while a higher-level expectation effect should be relatively indifferent to the location of the expected stimulus.

It is also possible that the variety of orientations presented in the prior-display attenuated the tilt aftereffect, but preserved a second manifestation of adaptation, which raises the contrast threshold required to detect a stimulus with the adaptor's orientation. While the tilt aftereffect is associated with changes in distribution of the population response of orientation-selective neurons, 'orientation-specific contrast adaptation' arises from a reduction in the sensitivity of the neurons specifically-tuned to the adaptor stimulus (Blakemore et al., 1973; He & MacLeod, 2001; Vergeer et al., 2018). The dominance of the unexpected stimulus may be explained by orientation-selective contrast adaptation reducing the effective contrast of the expected stimulus, as higher contrast stimuli dominate in binocular rivalry (Blake, 2001). For this reason, the expanded follow-up study of the tilt aftereffect should also include a task to specifically test for orientation-specific contrast adaptation.

Taken together, however, the prior-display may not have been the optimal strategy for instilling a sensory expectation. Another approach that would better divorce the role of expectation from adaptation is statistical learning. Here, participants are well-trained to associate pairs or sequences of stimuli within a continuous stream (Schapiro & Turk-Browne, 2015). Using complex stimuli, like natural images, in a statistical learning paradigm would address any concerns about adaptation arising from repeated exposure immediately prior to rivalry. The stream of images, presented binocularly, can then be interrupted by surreptitious rivalry probes containing an image predicted by the preceding sequence and an equally familiar image that is not expected to occur at that position in the sequence. By presenting probes on both binocularly-consistent and rivalrous frames of the stimulus stream, participants will not be alerted to the presence of competing stimuli on probe trials, minimising the risk of attentional confounds, exploratory eye movements etc. This approach has only been taken by one other binocular rivalry study (Denison et al., 2016), which produced the same result as the one reported in Chapter 2: unexpected images dominated rivalry.

Another limitation of this study is that the participants were completely inexperienced with binocular rivalry. It is quite possible that there were substantial inter-individual differences in their perceptual experience and in their internal criteria for the reporting of that experience. These factors may also have evolved at different rates

across participants as they acclimatised to rivalry over the course of the two testing sessions, although an analysis showed that there was no overall change in the rivalry dynamic across the trials within a session. It is also possible that internal expectations about the rhythm of alternations, led them to differentially attend to the stimulus they expected to dominate in the next transition. While the alternations in binocular rivalry cannot be completely controlled, attention can be used to influence the timing of the alternations (Chong & Blake, 2006; Hancock & Andrews, 2007). Future studies may benefit from providing training to participants with neutral stimuli to help to stablise these factors in advance of the testing sessions. For example, requiring participants to exhibit consistent mean dominance durations and alternation rates across several rivalry runs. Eye tracking could also be used in training to provide participants with live feedback on their fixation throughout the rivalry phase.

Discussions of the promise of predictive processing as a unifying framework of neural function often cite its neat accommodation of binocular rivalry. While the characterisation of rivalry as a hierarchical process is consistent with recent neurophysiological evidence, this study suggests that the predictive processing model of binocular rivalry may not be a sufficient account of a phenomenon that has eluded efforts to obtain a unifying explanation for centuries. Notwithstanding its limitations, the experiment reported in Chapter 2 joins a small number of studies that indicate that surprising events may immediately capture perception in binocular rivalry (Chopin & Mamassian, 2012; Denison et al., 2016). These studies contrast with others that report a facilitatory effect of expectation in rivalry (Andermane et al., 2020; Conrad et al., 2010; Denison et al., 2011). A candidate explanation for this schism is the temporal extent of expectation; the studies suggesting that perception is attracted to the expected stimulus tend to set expectations with the immediately preceding stimuli, while those showing the opposite effect tend to set expectations over longer stimulus sequences. Further work will be needed to explore the distinctions between priming, immediate expectation, and expectation based on longer-term statistical regularities.

Chapter 3. The Role of Expectation in Early Sensory Processing

As a computational scheme, predictive processing stands in sharp contrast to the traditional model of visual processing. In place of a feedforward system populated by feature-detectors, predictive processing models cast sensory input in a supervisory role, constraining the inferences of a cortical machinery that drives perception 'from the top down' (Clark, 2013). However, the comparison sometimes promotes a mischaracterisation of the classical model as a rigid process driven purely by feedforward sensory streams. Indeed, the traditional model does focus on feedforward processing, but this is more a reflection of the state of knowledge about the visual system than a dismissal of the relevance of sensory feedback. It was understood that early efforts to parse the computational organisation of sensory systems would likely be futile if the nature of information processing was not initially simplified to a unidirectional flow, given that the process appeared to be primarily driven by a feedforward pipeline (Stevens, 2012). By reducing cortical operations to a step-by-step

analysis of sensory data and neglecting the influence of backwards pathways, a basic schematic of the computational architecture became a more feasible goal for cognitive science. However, there is no contingent of cognitive scientists who deny that feedback from higher-levels can shape processing at lower levels (Pylyshyn, 1999). In fact, there is mounting evidence that building a stable perceptual representation of the visual scene requires feedback to help narrow the perceptual interpretations of sensory causes (de Lange et al., 2018; Gilbert & Li, 2013; Teufel & Fletcher, 2020). Feedback connections are thought to facilitate a variety of sensory functions, including figure-ground segregation, contour integration, and shape perception (Teufel & Nanay, 2017).

The pragmatic focus on bottom-up processing might also suggest that the traditional model does not provide for the influence of priors in sensory processing. On the contrary, certain priors are assumed to be embodied in the structure of the pathways supporting the operations of feedforward schemes, such that information processing is inherently constrained by persistent regularities in the sensory environment (Geisler, 2008; Girshick et al., 2011; Simoncelli, 2003; Simoncelli & Olshausen, 2001). More immediate contextual modulation is also accommodated in traditional approaches. For example, the responses of V1 neurons can be modulated by stimuli that lie beyond their receptive field, leading to suppressed responses when sensory input is consistent with a coherent structure in the broader visual field. This extra-classical receptive field effect can be seen as evidence that contextual information, available in the larger receptive fields of higher-level units, is used to suppress predictable input to V1 neurons (Rao & Ballard, 1999), but this effect can also be explained by processing schemes that eliminate redundancy and maximise the independence of neighbouring feedforward sensory channels, like divisive normalisation (Carandini & Heeger, 2011) or sparse coding models (Zhu & Rozell, 2013). Similarly, the phenomenon of repetition suppression, where repeated presentations of a stimulus lead to attenuated responses, is neatly explained by a system in which predictable inputs are suppressed (Hohwy, 2013), but it has proven remarkably difficult to separate accounts of repetition suppression based on top-down expectation from those based on neural adaptation (Grill-Spector et al., 2006; Grotheer & Kovacs, 2016; Walsh et al., 2020). Contemporary conceptions of the role of backwards channels in sensory processing include not just low-level contextual modulation and the allocation of attentional resources, but also the integration of sensory expectation (Gilbert & Li, 2013; Teufel & Nanay, 2017). Certain kinds of sensory prediction are relatively uncontroversial; for example, without corollary discharge predicting the translational consequences of the three saccades we make every second, perceptual experience would be beset by a nauseating world in perpetual motion (Wurtz & Sommer, 2004). The role of higher-level expectations, like expecting to hear Beethoven or see a python, is more controversial. However, this controversy does not stem from the exile of all such influences from the traditional model, rather it centres around the 'cognitive penetrability' of sensory processing (Lupyan, 2015; Pylyshyn, 1999; Teufel & Nanay, 2017); in other words, do these kinds of expectations alter the basic encoding of sensory information at the very earliest levels of the processing stream or do they exert some supplementary influence after a protected process of signal extraction?

So despite the fundamental differences in how predictive processing and traditional models characterise sensory processing, there is substantial overlap in the neural phenomena they can accommodate (Cao, 2020;

Heeger, 2017) and this profoundly complicates efforts to dissociate these models empirically. Chapter 3 pursued the divergence in what the models suggest should happen when incoming sensory information is confronted with sensory expectations in the early stages of sensory processing. By indexing early sensory processing with an SSVEP, this study was positioned to make a unique contribution to this literature. Most of the previous studies of expectation effects in early visual areas recorded BOLD activity (Feuerriegel et al., 2021a), but BOLD activity cannot segregate the responses of the expectation units and error units proposed to implement prediction error minimisation. In contrast, the superficial pyramidal cells that are primarily responsible for generating EEG signals (Cohen, 2017) have been proposed as the primary candidates to represent error units (Friston, 2009). Furthermore, the SSVEP is driven by sensory neurons' responses to the oscillating stimulus itself, so the signal arises from neurons tuned to the features of the stimulus. Therefore, when an unexpected stimulus is presented the response of the putative error units tuned to the unexpected features should be elevated, increasing the amplitude of the SSVEP. Conversely, when an expected stimulus is presented, the response of error units should be suppressed, reducing the amplitude of the SSVEP.

However, the analysis of the SSVEP indicated that there was no difference in the sensory response to expected and unexpected stimuli. Previous research indicated that attentional engagement with the stimulus may interact with the effects of expectation (Kok et al., 2012; Richter et al., 2019; Solomon et al., 2021; St. John-Saaltink et al., 2015), complicating the interpretation of the results, but equivalent responses were observed for expected and unexpected stimuli across two experiments that either directed the participants' attention to the fixation point or to the checkerboard stimulus itself. Another concern in interpreting these results is that participants may have failed to adequately encode the pattern of contrast changes presented in the practice block, which would explain the absence of a sensory expectation effect. However, analysis of the ERPs elicited by the final contrast element showed a clear vMMN. The vMMN demonstrates not only that participants encoded the regularity in the practice block, but also that this shaped sensory responses at some stage of the processing stream. It is difficult to determine the precise source of electrophysiological signals, but the vMMN is believed to originate in extrastriate areas (Kimura et al., 2012), while primary visual cortex is thought to generate the SSVEP in response to simple contrast stimuli like the checkerboard (Di Russo et al., 2007; Lauritzen et al., 2010; Vanegas et al., 2013). This arrangement was consistent with the more anterior topography of the vMMN and the earlier response of the SSVEP to the final contrast change (Figures 3.4 & 3.5). Together, this suggests that expectation plays a role in sensory processing, but may only exert an influence after an initial feedforward sweep, consistent with some alternative theoretical accounts of sensory expectation (Press et al., 2020; Trapp & Bar, 2015).

An important limitation in this study is that the SSVEP did not express a simple linear relationship with the contrast of the physical stimulus, particularly in the experiment where participants directly attended to the stimulus. Although there are clear responses to each contrast change, the amplitude of the signal did not scale in a straightforward way with the different contrast levels in the sequence. This was particularly surprising because several studies have demonstrated the excellent contrast sensitivity of the SSVEP (O'Connell et al., 2012; Steinemann et al., 2018; reviewed by Norcia et al., 2015). Furthermore, the SSVEP used in Chapters 5 and

6 provided an unambiguous index of the stimulus contrast, clearly segregating the higher-contrast target from the lower-contrast non-target stimulus (Figure 5.11.A). One key difference between these studies was the technique used to elicit the SSVEP. The experiment reported in Chapter 3 provoked the oscillatory activity by flickering the stimulus in piecemeal segments (Figure 3.2), but the experiment reported in Chapter 5 and 6 flickered the phase of the grating while keeping the entire stimulus on screen. The piecemeal approach may have reduced the neural response by diminishing the global contrast of the stimulus at any given moment. This ambiguity could be addressed in a replication of this study using the phase flicker to elicit the SSVEP. The stimuli used in each task may have also been an important factor. In both experiments described in Chapter 3, the stimulus was presented for longer periods of time than the stimulus in Chapter 5 and 6. This may have induced contrast adaptation, which would account for the decreasing trend in the amplitude of the SSVEP across the trial. Additionally, the large changes in contrast at each step in the sequence may have driven visual evoked potentials. This additional noise in the immediate response to the contrast change may explain the dips in the normalised SSVEP amplitude at each transition (Figure 3.4).

Finally, although it was not significant, the numerical trend was towards the expected stimulus eliciting a greater response than the unexpected stimulus. This is contrary to the hypothesis that a prediction error response would produce greater responses to unexpected stimuli, but is consistent with the SSVEP result in Chapter 6, where there was a significantly reduced SSVEP amplitude for unexpected stimuli. While it is important to not project meaning onto non-significant results, this simply provides some indication that further work assessing the role of expectation in sensory processing using the SSVEP will likely require large samples and trial numbers per participant to detect any effects, but for the reasons outlined above, this approach may also provide a valuable new perspective on the topic.

Examining The Mechanisms Implementing Expectation in Perceptual Decisions

Cognitive neuroscientists attempting to understand how the brain makes perceptual decisions have two sources of data: behaviour and recorded neural activity. Sequential sampling models have provided compelling accounts for both in humans, monkeys, and rodents (Brody & Hanks, 2016; Gold & Shadlen, 2007; O'Connell & Kelly, 2021; Ratcliff et al., 2016). The models suggest that perceptual decision-making is accomplished by capturing fleeting sensory evidence in an integrated representation which is updated over time until a decision threshold is reached. Given their success, one may naturally assume that this is a general theory that applies to all perceptual decision-making, but viable alternative accounts exist. In Chapter 5, these accounts, which propose narrow windows of integration or no integration at all, were collectively referred to as 'extreme leak models'. Computational modelling studies have illustrated the difficulty arbitrating between integration and extreme leak models based on behavioural fits alone (Ditterich, 2006; Purcell & Palmeri et al., 2017; Stine et al., 2020) and neurophysiological studies have indicated that the reality may be more complex than the perfect integration originally proposed in the Drift Diffusion Model (Yates et al., 2017). The literature is difficult to reconcile; while many studies favour perfect integration (Brunton et al., 2013; Evans et al., 2017; Ratcliff et al., 2016; Ratcliff & Smith, 2004; Waskom & Kiani, 2018), many report some leak in the integration process (Ditterich, 2006; Kiani et al., 2008; Stine et al., 2020; Usher & McClelland, 2001), and others suggest that only the most recent evidence influences the decision (Carland et al., 2016; Cisek et al., 2009; Thura et al., 2012). For those attempting to develop standardised solutions to these problems, accounting for the role of noise is likely to be critical.

The primary advantage of evidence integration is that averaged samples flush noise out of the signal, enhancing perceptual sensitivity and decision-making efficiency. For this reason, the random dot motion stimulus, which embeds the motion signal amongst random noise, has been the battleground for the debate around integration. Critics contend that accurate estimates of the sensory evidence should emerge rapidly for tasks like this, where stimuli are typically presented at a static evidence level, because the noise minimisation achieved by averaging across samples quickly saturates, after which further samples are redundant (Carland et al., 2016). They argue that perfect integration would be wasteful and maladaptive for animals that require flexibility and sensitivity to novelty in dynamic environments (Glaze et al., 2015; Ossmy et al., 2013). Instead, some propose that perceptual decision-making is better accounted for by models that disregard sensory samples after about a hundred milliseconds (Cisek et al., 2009). In tasks where physical stimulus noise is absent, leaving only noise intrinsic to sensory processing itself, the proposed redundancy of integration is only exaggerated; so these tasks seem to be good candidates for exposing extreme leak strategies. Indeed, simple

perceptual detection tasks without any noise are traditionally explained by momentary evidence evaluation with schemes like probability summation or extrema detection (Watson, 1979). Very few studies have investigated perceptual decisions using stimuli with little physical noise. The study reported in Chapter 5 sought to determine if participants integrate evidence on low-noise tasks and provide some insight about the origins of the CPP.

Within the framework of the most prominent example of a sequential sampling model, the Drift Diffusion Model, two mechanisms can represent the influence of expectation on the decision process: a starting point bias and a drift bias. A starting point bias shifts the decision variable closer to the expected bound at evidence onset, while a drift bias prejudices the rate of evidence accumulation in favour of the expected outcome. The overwhelming majority of studies that have investigated how expectations influence perceptual decision-making have endorsed the starting point bias account, but two important studies have provided evidence that there may be more subtle dynamics at play (Hanks et al., 2011; Kelly et al., 2021). Both studies constrained model parameters with empirically-derived neural indices and both studies provided evidence that a drift bias is an important component of the decision adjustments associated with priors. This is a compelling indication that the emphasis on model parsimony may promote models that can capture the intricate interplay of multiple weak influences with a single dominant variable, like the starting point bias, suggesting that careful examination of neural signatures of decision-making may reveal hints of the influence of more subtle adjustments lurking in the decision process. Chapter 6 looked to capture neural signatures of decision adjustments representing the influence of prior expectation in perceptual decision-making.

Chapter 5. Evidence Accumulation in Perceptual Decisions About Low-Noise Stimuli

The literature discussed above illustrates the difficulty with definitively identifying integration based on behavioural modelling, so the study reported in Chapter 5 was specifically designed to test for signatures of integration with model-independent behavioural and neurophysiological indices. The study incorporated pulses of evidence into a contrast discrimination task to measure the longevity of the pulses' influence over the decision process. While extreme leak accounts suggest that evidence is only retained in the decision process for 100-250 ms, the 150 ms contrast pulses were found to continue to influence choice accuracy ~1000 ms after the pulse onset. This enduring influence was also reflected in neurophysiological indices of evidence accumulation and motor preparation for at least 500 ms after pulse onset. Furthermore, extreme leak accounts predict that evidence samples outside of the narrow integration window are discarded, but on 'gap pulse' trials, where all evidence was eliminated for 150 ms, the CPP showed no signs of a return towards baseline. Together, this study provides important evidence that participants continued to integrate sensory evidence, even when the stimulus contained little physical noise. In future work, it would be useful to simulate the responses of the Drift Diffusion Model and extreme leak models to further illustrate this point.

This raises the question of why an integration strategy is useful when there is little physical noise to eliminate? A promising explanation was hinted at in an analysis of the SSVEP. Before evidence onset, the marginal amplitude of the SSVEP signal was a significant predictor of reaction time, accuracy, and the SSVEP signal amplitude after evidence onset. As no evidence had been presented at this point in the trial, this suggests that internal noise produced differences in the SSVEP, which may also be an important factor in the evolution of the decision. The brain has no reason to discriminate between external and internal sources of noise because integrating evidence is an efficient strategy to minimise the influences of both. From this perspective, any difficult perceptual task would benefit from evidence accumulation, whether or not the stimulus itself was noisy. The more difficult the task, the smaller the ratio between the magnitude of internal noise and sensory evidence, and the greater the advantage of minimising that noise through integration. One caveat here is that this interpretation is based on the assumption that the differences in pre-evidence SSVEP amplitude reflect rapid fluctuations in the sensory representation and not sustained changes in the quality of the representation across the duration of the trial. If there was a sustained reduction in the marginal target SSVEP amplitude across the entire trial, it is not clear that integration would confer any advantages over other decision strategies. Future analyses can address this question by investigating if the volatility of the SSVEP signals during evidence presentation predicts reaction time and accuracy. This would confirm that the volatility of the recorded signal was linked to volatility of the sensory evidence and was not a product of measurement noise unrelated to the decision, and therefore, that internal noise was a factor in the decision process.

The CPP is believed to represent a stage of the decision process that is upstream of motor preparation (O'Connell & Kelly, 2021), but there is only one study providing direct evidence for this claim. Kelly and O'Connell (2013) measured the latency of the CPP and LRP's response to the onset of gradual and prolonged contrast targets and found that the CPP response emerged ~150 ms before the LRP response. In Chapter 5, this result was replicated by measuring the latency of the CPP and LRP response to the brief and discrete pulses of evidence. Examining the timing of the divergence of responses to pulses that increased or reversed the evidence provided a reliable timing estimate of the onset of the pulse response for each signal. In line with Kelly and O'Connell (2013), the analysis indicated that the CPP response preceded the LRP response by ~85 ms, reaffirming the characterisation of the CPP as a window onto an abstract stage of the decision process.

The discovery of the CPP was a turning point in research on decision-making in humans, but the neuronal origins of this signal are largely unknown. A final analysis sought to better characterise what exactly the CPP represents by measuring its response to counter-evidence. The CPP may represent cumulative evidence for a particular perceptual hypothesis, in which case counter-evidence should drive the signal back towards baseline. Alternatively, the CPP could arise from the coalesced responses of multiple accumulators simultaneously and independently representing the evidence for each of the decision alternatives, in which case counter-evidence should increase the amplitude of the global signal. There was an absolute increase in the CPP amplitude following both gap and reverse pulses on correct response trials. Although this is consistent with the multiple accumulators account, the absence of a pulse effect on error trials and the observation that there was no difference in the response to gap and reverse pulses on correct response trials are not as easily explained.

Future work simulating the response of different CPP constructions will provide valuable insight about exactly the pattern of responses to the pulses that could be expected under different conditions.

Chapter 6. The Mechanisms of Prior Influence in Perceptual Decision-Making

Having established that participants did accumulate evidence while performing this task in Chapter 5, the final empirical chapter sought to characterise the implementation of expectation in perceptual decisions with electrophysiological indices of three key stages of the decision process: sensory encoding, evidence accumulation, and response preparation. Specifically, it was hypothesised that a drift bias would be evident in the response to discrete pulses of evidence that could either be consistent with the expectation, conflict with the expectation, or remove all evidence. In addition, it has been suggested that the drift bias may be dynamic, changing the extent of the prior's influence over the duration of the trial (Deneve, 2012; Hanks et al., 2011). To test this idea, pulses of evidence were presented at unpredictable onset times in the first 500 ms of the trial.

Several studies have reported that expectations elicit preparatory sensory activity, resembling a predictive template of the upcoming stimulus (Blom et al., 2020; Chong et al., 2016; Kok et al., 2012, 2017). It is plausible that this preparatory activity represents a sensory-level source of either a starting point bias, if this predictive representation is itself accumulated as evidence (Feuerriegel et al., 2021b), or a drift bias, if this activity represents a readiness to sharpen the representation of upcoming expectation-congruent sensory input (Kok et al., 2012). There was no evidence that the cue evoked this kind of preparatory activity in the SSVEP during the baseline phase of the trial. However, there was a significant effect of the cue on the SSVEP amplitude during evidence presentation. The marginal amplitude of the SSVEP signal representing the target was smaller on invalid cue trials, the opposite pattern to what is typically shown in expectation research, where expected stimuli elicit suppressed responses. This appears to be a promising candidate for a sensory-level drift bias, since prior expectations are shaping the encoding of the sensory evidence itself, but two observations complicate this interpretation. First, the paradigm was specifically designed to detect such an effect based on the response to the pulses of evidence, but there was no effect of expectation on the SSVEP responses to the pulses. This might be explained by an additive drift bias, which enhances the baseline representation of the expected stimulus without changing the sensitivity to congruent/incongruent sensory input. Although there was no evidence of an additive bias in the pre-evidence SSVEP, the effect may be dependent on a certain level of stimulus engagement that was more reliably achieved after evidence onset. Second, when the data were separated into trials where the target was repeated from the previous trial and those where the target stimulus alternated, the SSVEP effect appeared to be driven primarily by repetition trials. This could reflect conflicting sources of expectation from the cue and choice history. The analyses did not directly address the influence of choice history in this study, but previous research has shown that choice history can bias the accumulation of evidence, potentially originating from a sensory-level modulation (Urai et al., 2019). Computational modelling of this dataset may help to resolve some of this ambiguity. If modelling suggests that there is a drift bias, a

follow-up experiment with a simplified paradigm and an attention manipulation (similar to that used in Chapter 3) may help to refine the characterisation of this effect. Alternatively, if modelling indicates that there is no drift bias, this effect may not be a critical component of the decision process in this study.

There was strong evidence that the cue induced a starting point bias at the motor level. A divergence in motor preparation favouring the cued response emerged shortly after cue presentation and was a pronounced effect well in advance of evidence onset (Figure 6.14.A). Consistent with previous research (de Lange et al., 2013), the magnitude of the preparatory bias predicted the timing and accuracy of the forthcoming decision. Indeed, analyses of the motor preparation signals after evidence onset suggested that this preparatory activity had a lasting influence on the evolution of the decision, rigging the starting conditions to expedite validly-cued correct responses and forcing evidence accumulation to overcome an initial deficit to reach the correct bound on invalid cue trials. This latter effect was evident in the CPP amplitude at response, indicating greater evidence accumulation was necessary on trials where the response contradicted the cue.

There was mixed evidence that a subtle drift bias may also have shaped the decision. The influence of a starting point bias is expected to be most evident at the leading edge of a reaction time distribution and cease to be a force in the decision for later reaction times, but the cue effect on reaction time could be seen across the full reaction time distribution for both correct and error responses (Figure 6.3.C). Similarly, there was an enduring influence of the cue on the conditional accuracy functions, indicating that expectation continued to affect the decision even for the slowest responses (Figure 6.8.A). It was hoped that a prominent and discrete burst of evidence would act as a prism for an evidence-dependent drift bias, refracting decision signals to maximise any divergence in the effects of that evidence in different expectational states and different phases of the decision. However, the CPP and behavioural analyses of the pulse effects provided no evidence of an expectation modulation, which is consistent with the previous suggestion that an additive drift bias may additively influence the decision process independent of the sensory evidence. Finally, there was no cue-effect in the response-locked slope of the CPP, but it is difficult to interpret this result without modelling because it is very likely the product of a complicated interaction between several factors of varying influence. One of these complicating factors is dynamic urgency, which was apparent in the early onset of motor preparation in this study. Kelly et al. (2021) provided a clear demonstration of how easy it would be to overlook a drift bias in the analysis of the neurophysiological and behavioural data alone amidst all of these competing influences, but they highlighted the value of computational modelling in attempting to parse this complexity. It is intended to implement their neurally-informed modelling approach in a second stage of analysis of this dataset in the near future.

Summary

Few topics in cognitive neuroscience are as widely accepted and deeply controversial as the role of prior experience in perceptual processing. As a response to the indeterminate sensory world, prior experience is the celebrated Bayesian response; as an inspiration for cortical architecture, phylogenetic priors are indisputable

(Ganguli & Simoncelli, 2014; Simoncelli, 2003); experience is commonly regarded as the organising force in the ontogeny of perceptual systems (Flom, 2014; Lewkowicz, 2014), beginning in utero (e.g. De Casper et al., 1994); sensitivity to the history of stimulation in fundamental neural mechanisms like adaptation are beyond question (Solomon & Kohn, 2014; Vogels, 2016); and prediction is a central component in theories of several neural processes, like motor control (McNamee & Wolpert, 2019). Yet the possibility that priors dynamically shape sensory encoding itself was somewhat neglected in an era of colossal achievements by those charting the frontiers of a feedforward model of sensory processing. In the last two decades, there has been renewed interest in exploring the purpose of feedback projections that saturate early sensory regions (Briggs, 2020) and probing the dogma of cognitive penetrability (Bubic et al., 2010; Lupyan, 2015; Pouget et al., 2013; Teufel & Nanay, 2020). The work presented in this thesis looked to contribute to this discussion with investigations of the role of priors in basic sensory processing, perceptual decision-making, and selecting the contents of perceptual awareness. The results suggest that expectation can manifest in unexpected ways and many of these observations point to promising avenues for further investigation. It is hoped that future work can offer some elaboration.

References

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience*, *7*, 1057–1058.

Almasi, A., Meffin, H., Cloherty, S. L., Wong, Y., Yunzab, M., & Ibbotson, M. R. (2020). Mechanisms of feature selectivity and invariance in primary visual cortex. *Cerebral Cortex*, *30*, 5067–5087.

Andermane, N., Bosten, J., Seth, A., & Ward, J. (2020). Individual differences in the tendency to see the expected.

Baillargeon, R. (2004). Infants' reasoning about hidden objects: evidence for event-general and event-specific expectations. *Developmental Science*, *7*, 391–414.

Bar, M., & Ullman, S. (1996). Spatial context in recognition. Perception, 25, 343–352.

Bar, M. (2004). Visual objects in context. Nature Reviews. Neuroscience, 5, 617–629.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1(01).

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology?. *Perception*, 1(4), 371-394.

Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, *331*, 83–87.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143–177.

Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*, 77–80.

Blakemore, C., Muncey, J. P., & Ridley, R. M. (1973). Stimulus specificity in the human visual system. *Vision Research*, *13*, 1915–1931.

Blake, R. (2001). A Primer on Binocular Rivalry, Including Current Controversies. *Brain & Mind*, 5–38.

Blom, T., Feuerriegel, D., Johnson, P., Bode, S., & Hogendoorn, H. (2020). Predictions drive neural representations of visual events ahead of incoming sensory information. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 7510–7515.

Bond, A. H. (2004). An information-processing analysis of the functional architecture of the primate neocortex. *Journal of Theoretical Biology*, *227*, 51–79.

Briggs, F. (2020). Role of feedback connections in central visual processing. *Annual Review of Vision Science*, *6*, 313–334.

Brody, C. D., & Hanks, T. D. (2016). Neural underpinnings of the evidence accumulator. *Current Opinion in Neurobiology*, *37*, 149–157.

Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science*, *340*, 95–98.

Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, *4*, 25.

Cao, R. (2020). New labels for old ideas: predictive processing and the interpretation of neural signals. *Review of Philosophy and Psychology*, *11*, 517–546.

Carandini, M., & Heeger, D. J. (2011). Normalization as a canonical neural computation. *Nature Reviews. Neuroscience*, *13*, 51–62.

Carland, M. A., Marcos, E., Thura, D., & Cisek, P. (2016). Evidence against perfect integration of sensory information during perceptual decision making. *Journal of Neurophysiology*, *115*, 915–930.

Chong, E., Familiar, A. M., & Shim, W. M. (2016). Reconstructing representations of dynamic visual objects in early visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 1453–1458.

Chong, S. C., & Blake, R. (2006). Exogenous attention and endogenous attention influence initial dominance in binocular rivalry. *Vision Research*, *46*, 1794–1803.

Chopin, A., & Mamassian, P. (2012). Predictive properties of visual adaptation. *Current Biology*, *22*, 622–626.

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: the urgency-gating model. *The Journal of Neuroscience*, *29*, 11560–11571.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204.

Cohen, M. X. (2017). Where does EEG come from and what does it mean? *Trends in Neurosciences*, *40*, 208–218.

Conrad, V., Bartels, A., Kleiner, M., & Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision*, *10*, 27.

Daw, N. W. (2009). The foundations of development and deprivation in the visual system. *The Journal of Physiology*, *587*, 2769–2773.

DeCasper, A. J., Lecanuet, J.-P., Busnel, M.-C., Granier-Deferre, C., & Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. *Infant Behavior and Development*, *17*, 159–164.

Deneve, S. (2012). Making decisions with unknown sensory reliability. *Frontiers in Neuroscience*, *6*, 75.

Denison, R. N., Piazza, E. A., & Silver, M. A. (2011). Predictive Context Influences Perceptual Selection during Binocular Rivalry. *Frontiers in Human Neuroscience*, *5*, 166.

Denison, R. N., Sheynin, J., & Silver, M. A. (2016). Perceptual suppression of predicted natural images. *Journal of Vision*, *16*, 6.

den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*, 548.

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*, 764–779.

de Lange, F. P., Rahnev, D. A., Donner, T. H., & Lau, H. (2013). Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *The Journal of Neuroscience*, *33*, 1400–1410.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*, 415–434.

Ditterich, J. (2006). Stochastic models of decisions about motion direction: behavior and physiology. *Neural Networks*, *19*, 981–1012.

Di Russo, F., Pitzalis, S., Aprile, T., Spitoni, G., Patria, F., Stella, A., … Hillyard, S. A. (2007). Spatiotemporal analysis of the cortical sources of the steady-state visual evoked potential. *Human Brain Mapping*, *28*, 323–334.

Evans, N. J., Hawkins, G. E., Boehm, U., Wagenmakers, E.-J., & Brown, S. D. (2017). The computations that support simple decision-making: A comparison between the diffusion and urgency-gating models. *Scientific Reports*, *7*, 16433.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47.

Feuerriegel, D., Blom, T., & Hogendoorn, H. (2021b). Predictive activation of sensory representations as a source of evidence in perceptual decision-making. *Cortex*, *136*, 140–146.

Feuerriegel, D., Vogels, R., & Kovács, G. (2021a). Evaluating the evidence for expectation suppression in the visual system. *Neuroscience and Biobehavioral Reviews*, *126*, 368–381.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 15822–15826.

Flom, R. (2014). Perceptual narrowing: retrospect and prospect. *Developmental Psychobiology*, *56*, 1442–1453.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*, 815–836.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*, 293–301.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*, 127–138.

Furmanski, C. S., & Engel, S. A. (2000). An oblique effect in human primary visual cortex. *Nature Neuroscience*, *3*, 535–536.

Ganguli, D., & Simoncelli, E. P. (2014). Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural Computation*, *26*, 2103–2134.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*, 167–192.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews. Neuroscience*, *14*, 350–363.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*, 926–932.

Glaze, C. M., Kable, J. W., & Gold, J. I. (2015). Normative evidence accumulation in unpredictable environments. ELife, 4.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

Gold, J. I., & Stocker, A. A. (2017). Visual Decision-Making in an Uncertain and Dynamic World. *Annual Review of Vision Science*, *3*, 227–250.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*, 14–23.

Grotheer, M., & Kovács, G. (2016). Can predictive coding explain repetition suppression? *Cortex*, *80*, 113–124.

Hancock, S., & Andrews, T. J. (2007). The role of voluntary and involuntary attention in selecting perceptual dominance during binocular rivalry. *Perception*, *36*, 288–298.

Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *The Journal of Neuroscience*, *31*, 6339–6352.

Heeger, D. J. (2017). Theory of cortical function. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 1773–1782.

He, S., & MacLeod, D. I. (2001). Orientation-selective adaptation and tilt after-effect from invisible patterns. *Nature*, *411*, 473–476.

Hermundstad, A. M., Briguglio, J. J., Conte, M. M., Victor, J. D., Balasubramanian, V., & Tkačik, G. (2014). Variance predicts salience in central sensory processing. Elife, 3, e03722.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition*, *108*, 687–701.

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*, 106–154.

Kelly, S. P., Corbett, E. A., & O'Connell, R. G. (2021). Neurocomputational mechanisms of prior-informed perceptual decision-making in humans. *Nature Human Behaviour*, *5*, 467–481.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *The Journal of Neuroscience*, *33*, 19434–19441.

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *The Journal of Neuroscience*, *28*, 3017–3029.

Kimura, M. (2012). Visual mismatch negativity and unintentional temporal-context-based prediction in vision. *International Journal of Psychophysiology*, *83*, 144–155.

Knapen, T., Rolfs, M., Wexler, M., & Cavanagh, P. (2010). The reference frame of the tilt aftereffect. Journal of Vision, 10.

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, *75*, 265–270.

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 10473–10478.

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex*, *22*, 2197–2206.

Kording, K. P. (2014). Bayesian statistics: relevant for the brain? *Current Opinion in Neurobiology*, *25*, 130–133.

Lauritzen, T. Z., Ales, J. M., & Wade, A. R. (2010). The effects of visuospatial attention measured across visual cortex using source-imaged, steady-state EEG. Journal of Vision, 10.

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology*, *56*, 292–315.

Li, B., Peterson, M. R., & Freeman, R. D. (2003). Oblique effect: a neural basis in the visual cortex. *Journal of Neurophysiology*, *90*, 204–217.

Lochmann, T., & Deneve, S. (2011). Neural processing as causal inference. *Current Opinion in Neurobiology*, *21*, 774–781.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.

Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: predictive systems are penetrable systems. *Review of Philosophy and Psychology*, *6*, 547–569.

Maniscalco, B., Lee, J. L., Abry, P., Lin, A., Holroyd, T., & He, B. J. (2018). Neural integration of stimulus history underlies prediction for naturalistically evolving sequences. *The Journal of Neuroscience*, *38*, 1541–1557.

Mathôt, S., & Theeuwes, J. (2013). A reinvestigation of the reference frame of the tilt-adaptation aftereffect. *Scientific Reports*, *3*, 1152.

McNamee, D., & Wolpert, D. M. (2019). Internal models in biological control. *Annual Review of Control, Robotics, and Autonomous Systems*, *2*, 339–364.

Mitchell, D. E., Freeman, R. D., & Westheimer, G. (1967). Effect of orientation on the modulation sensitivity for interference fringes on the retina. *Journal of the Optical Society of America*, *57*, 246–249.

Nassi, J. J., & Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nature Reviews. Neuroscience*, *10*, 360–372.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*, 4.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*, 1729–1735.

O'Connell, R. G., & Kelly, S. P. (2021). Neurophysiology of Human Perceptual Decision-Making. *Annual Review of Neuroscience*, *44*, 495–516.

Ossmy, O., Moran, R., Pfeffer, T., Tsetsos, K., Usher, M., & Donner, T. H. (2013). The timescale of perceptual evidence integration can be adapted to the environment. *Current Biology*, *23*, 981–986.

Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*, 519–526.

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, *16*, 1170–1178.

Press, C., Kok, P., & Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences*, *24*, 13–24.

Purcell, B. A., & Palmeri, T. J. (2017). Relating accumulator model parameters and neural dynamics. *Journal of Mathematical Psychology*, *76*, 156–171.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*, 341–365; discussion 366.

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

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: current issues and history. *Trends in Cognitive Sciences*, *20*, 260–281.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367.

Richter, D., & de Lange, F. P. (2019). Statistical learning attenuates visual activity only for attended stimuli. ELife, 8.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*, 1019–1025.

Russell, B. (2001). The problems of philosophy (2nd ed.). Oxford University Press.

Rust, N. C., & Palmer, S. E. (2021). Remembering the past to see the future. *Annual Review of Vision Science*, *7*, 349–365.

Schapiro, A., & Turk-Browne, N. (2015). Statistical Learning. In *Brain Mapping* (pp. 501–506). Elsevier.

Seriès, P., & Seitz, A. R. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience*, *7*, 668.

Serre, T. (2015). Hierarchical models of the visual system. In D. Jaeger & R. Jung (Eds.), *Encyclopedia of computational neuroscience* (pp. 1309–1318). New York, NY: Springer New York.

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. Current Opinion in Neurobiology, 13, 144–149.

Simoncelli, E P, & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216.

Solomon, S. G., & Kohn, A. (2014). *Moving Sensory Adaptation beyond Review Suppressive Effects in Single Neurons*. *24*, R1012–R1022.

Solomon, S. S., Tang, H., Sussman, E., & Kohn, A. (2021). Limited evidence for sensory prediction error responses in visual cortex of macaques and humans. *Cerebral Cortex*, *31*, 3136–3152.

Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, *144*, 1089–1104.

Steinemann, N. A., O'Connell, R. G., & Kelly, S. P. (2018). Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nature Communications*, *9*, 3627.

Stevens, K. A. (2012). The vision of David Marr. *Perception*, *41*, 1061–1072.

Stine, G. M., Zylberberg, A., Ditterich, J., & Shadlen, M. N. (2020). Differentiating between integration and non-integration strategies in perceptual decision making. ELife, 9.

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*, 578–585.

St John-Saaltink, E., Utzerath, C., Kok, P., Lau, H. C., & de Lange, F. P. (2015). Expectation suppression in early visual cortex depends on task set. *Plos One*, *10*, e0131172.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews. Neuroscience*, *15*, 745–756.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*, 403–409.

Teufel, C., & Fletcher, P. C. (2020). Forms of prediction in the nervous system. *Nature Reviews. Neuroscience*, *21*, 231–242.

Teufel, C., & Nanay, B. (2017). How to (and how not to) think about top-down influences on visual perception. *Consciousness and Cognition*, *47*, 17–25.

Thura, D., Beauregard-Racine, J., Fradet, C.-W., & Cisek, P. (2012). Decision making by urgency gating: theory and experimental support. *Journal of Neurophysiology*, *108*, 2912–2930.

Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, *1339*, 190–198.

Trenholm, S., & Krishnaswamy, A. (2020). An Annotated Journey through Modern Visual Neuroscience. *The Journal of Neuroscience*, *40*, 44–53.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*, 1934–1945.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *The Journal of Neuroscience*, *30*, 11177–11187.

Urai, A. E., de Gee, J. W., Tsetsos, K., & Donner, T. H. (2019). Choice history biases subsequent evidence accumulation. ELife, 8.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.

Vanegas, M. I., Blangero, A., & Kelly, S. P. (2013). Exploiting individual primary visual cortex geometry to boost steady state visual evoked potentials. *Journal of Neural Engineering*, *10*, 036003.

Van Essen, David C., & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, *6*, 370–375.

Van Essen, D C, & Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron*, *13*, 1–10.

Vergeer, M., Mesik, J., Baek, Y., Wilmerding, K., & Engel, S. A. (2018). Orientation-selective contrast adaptation measured with SSVEP. *Journal of Vision*, *18*, 2.

Vogels, R. (2016). Sources of adaptation of inferior temporal cortical responses. *Cortex*, *80*, 185–195.

Wade, N. J. (2021). Helmholtz at 200. *I-Perception*, *12*, 20416695211022376.

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*, 242–268.

Waskom, M. L., & Kiani, R. (2018). Decision Making through Integration of Sensory Evidence at Prolonged Timescales. *Current Biology*, *28*, 3850-3856.e9.

Watson, A. B. (1979). Probability summation over time. *Vision Research*, *19*, 515–522.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*, 598–604.

Wiesel, T. N., & Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology*, *26*, 1003–1017.

Wiese, W., & Metzinger, T. (2017). Vanilla PP for Philosophers: A Primer on Predictive Processing. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing* (pp. 1–18). Frankfurt am Main.

Wurtz, R. H., & Sommer, M. A. (2004). Identifying corollary discharges for movement in the primate brain. *Progress in Brain Research*, *144*, 47–60.

Yates, J. L., Park, I. M., Katz, L. N., Pillow, J. W., & Huk, A. C. (2017). Functional dissection of signal and noise in MT and LIP during decision-making. *Nature Neuroscience*, *20*, 1285–1292.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, *24*, 667–677.

Zhu, M., & Rozell, C. J. (2013). Visual nonclassical receptive field effects emerge from sparse coding in a dynamical system. *PLoS Computational Biology*, *9*, e1003191.