Dissociable Sources of Uncertainty in Perceptual Decision Making

Elizabeth Michael



Department of Experimental Psychology & Magdalen College
DPhil, Experimental Psychology
University of Oxford
Hilary Term 2015

Table of Contents

Abstract i

Acknowledgements ii

Long Abstract iii

Chapter 1 1

1.1. Hierarchical organisation is a key principle of the visual system. 1

1.2. Theoretical models of perceptual categorisation - optimal observers and mechanistic implementations. 4

1.3. Comparison of Theoretical Model Predictions and Functioning of Neural Systems. 14

1.4. Neural studies of categorisation have demonstrated the distributed nature of the representations necessary for categorisation. 18

1.5. Human neuroimaging studies have found several neural correlates for decision variables, and have highlighted the interconnected cognitive processes underlying accurate perceptual choice. 21

1.6. Cognitive theories of categorisation have argued for several different forms of categorical representation. 25

1.7. Chapter Summary 29

Chapter 2 30

2.1. Perceptual averaging in human observers: the global level of perception as a statistical summary measure. 30

2.2. Explicit report of the summary measure in the absence of such knowledge about the component features. 34

2.3. Representation of summary statistics can be partially dissociated from representations reliant on selective attention. 37

2.4. Perceptual summary representation reflects real averaging strategies but does not necessarily rely on equally weighted items. 40

2.5. Summary statistics act as cues for identity in noisy or low-resolution environments. 44

2.6. To what extent is a full summary stimulus distribution represented in the brain? 46

2.7. Chapter Summary 49

Chapter 3 - fMRI 1 50

3.1. Methods. 55
3.2. Results. 63
3.3. Discussion. 78

Chapter 4 - EEG 82

4.1. Methods. 85
4.2. Results. 90
4.3. Discussion. 104

Chapter 5 - Behavioural 107

5.1. Methods. 111
5.2. Results. 115
5.3. Discussion. 129

Chapter 6 -fMRI 2 132

6.1. Methods. 136
6.2. Results. 141
6.3. Discussion. 154

Chapter 7 158

7.1. Dissociable sources of perceptual uncertainty. 158
7.2. Relationship to Summary Representation. 161
7.3. How might mean and variability be represented by the neural code? 163
7.4. Summary values as cues for adaptation. 164
7.5. Final Conclusions 166

Appendices

Appendix 1 168

Figure A1.1-A1.2 168
Tables A1.1-A1.2 171

Appendix 2 177

Tables A2.1 -A2.2 177

Appendix 3 179

Fig A3.1 179
Table A3.1 - A3.5 180

Appendix 4 186

Tables A4.1.1-A4.3 186

References 190

Dissociable sources of uncertainty in perceptual decision making

Elizabeth Michael

Department of Experimental Psychology & Magdalen College, Oxford
Hilary Term 2015

The natural world provides sensory systems with noisy and ambiguous information, which is often transformed into a more stable categorical percept. This thesis aims to investigate the nature of the neural representations in the visual system that support this transformation. To do so, we will employ a behavioural task that requires participants to average several independent sources of perceptual information. This task allows for the dissociation of two theoretically orthogonal sources of decision uncertainty: the mean distance of the perceptual information from a category boundary and the variability of the evidence under consideration. Behaviourally, both decreasing the mean distance to bound of information and increasing information variability are associated with increased errors and prolonged response times. We will present a computational model that can account for the independent behavioural effects of these two sources of uncertainty by assuming that categorical decisions are made on the basis of a probabilistic transformation of perceptual evidence. BOLD measurements demonstrate that these effects of mean and variability are supported by a partially dissociable network of brain regions. Electroencephalography demonstrates the differential influence of mean and variance in the pre- and post-decision period. Furthermore, we show that there is adaptation at the level of the perceptual representation to the information variance. Not only does this show that the visual system must represent information at the summary level, in addition to individual feature-based representation, but it also suggests that the costs associated with this form of perceptual uncertainty can be largely mitigated by the adoption of a more suitable representational range.

This work was funded by a 3-year studentship from the Medical Research Council, and was conducted under the supervision of Dr Christopher Summerfield.

Acknowledgements

Firstly, I would like to thank the Department of Experimental Psychology and Magdalen College for providing an outstanding environment in which to work, learn and live. I have had the opportunity to interact with so many wonderful people on a daily basis and have learnt more from this than any thesis could ever describe.

Primarily I owe a huge debt of gratitude to everyone in the Summerfield lab, past and present. The lab has grown a little since I first arrived, and it just keeps on getting better! I would especially like to thank Vincent, founding member of Team Squircle, for all his help and advice in the early months. Thanks to those at FMRIB and in Granada who sat through (and occasionally participated in) a few long weekends of data collection. The ACC Lab has been a constant source of invaluable information along the way. For hours of much-needed help with EEG I would like to thank Marike and Annika (and also for the knitting, wine and support). Any ideas get more interesting when there's someone around to talk about them with, so, Jan, please never stop asking all those questions.

To all my friends, thank you for keeping the faith (even when there was no reply). I would particularly like to thank H. Chilswell and H. Alan (and associated members) for the dinners, experimental baking, Contract Law, tolerance, advice, jelly, House Rules, incredible accounting skills, DIY knowledge and so much more.

To my family, thank you for your love and understanding. To Paul, Caroline and Lily, for all the Happy Days. To my parents, for encouraging me to make my own decisions. Mum - thank you for keeping us all together and for all that you do.

And finally - Chris. It feels quite some time has passed since I first sat in your office to talk about potential projects. Not all of this time has been stress-free, but I have always known that I was in the right place. You have invested so much, in both me and my work. I will always be grateful that you took a bit of a risk on a girl who didn't really say anything - after all this time I am starting to find my own voice and this is largely down to you. Thank you.

Long Abstract

Chapter 1 presents a view of the visual system as a hierarchically organised series of increasingly complex stages of information transformation. Theoretical models of perceptual categorisation have provided a clear framework for understanding these transformations, in which a noisy and continuous sensory input is transformed into a more stable conceptual representation. More recent neuroscientific studies have aimed to identify the neural correlates of this process, and converging evidence from single unit and neuroimaging studies suggest that perceptual decisions rely on the integration of sensory evidence from early sensory cortices. A reliable network of brain regions beyond sensory cortex responds to the uncertainty associated with the decision-relevant evidence, including parietal, lateral prefrontal and medial prefrontal cortices. However, the exact nature of the representations supported by these brain regions remains elusive.

Chapter 2 introduces an alternative view of perceptual representation, in which summary or "ensemble" representations of sensory information are afforded a privileged role. This view encompasses previous ideas such as global/local distinctions, or "gist"-based vs feature-based modes of representation. More recent work suggests that this summary level of representation directly encodes the statistical quantities associated with visual input, with a surprising level of fidelity. Specifically, several studies have demonstrated the ability of observers to report the average feature across a set, in the absence of any demonstrable knowledge of the individual component features. In this thesis, we will present evidence to support the claim that human observers are sensitive not only to the average, but also to the variance of visual input. In a decision-making framework, these quantities also reflect two theoretically dissociable sources of uncertainty. We present a task that is designed to orthogonalise the mean and variability and measure their independent contributions to perceptual choice.

Chapter 3 presents an fMRI experiment in which we present neural dissociations between two sources of decision-relevant perceptual uncertainty (mean and variability of feature information). This supports behavioural findings of independent contributions of mean and variability to categorisation performance, in that both decreasing mean feature distance to boundary and increasing variability are associated with decreased accuracy and increased response latencies. A model (the log-probability, or LPR, model) is presented which explains both the effects of mean and variance by assuming that the feature-level information is transformed into a probabilistic space ("decision space") before being averaged. This model gains support here from the neural data, as the BOLD signal is best explained by uncertainty in decision space, as opposed to the original sensory space.

Although many regions show a positive correlation between BOLD signal and decision uncertainty, medial prefrontal cortex (mPFC) showed a unique response such that it scaled positively with the mean but negatively with the variability, a response profile not predicted by several prominent theories of mPFC function. We discuss possible alternative models, and suggest that the BOLD signal in this region may reflect a gain control signal applied to neural representations falling close to the category boundary, an explanation consistent with several features of the LPR model.

Chapter 4 aims to expand on the fMRI findings presented in Chapter 3, by exploring whether the observed pattern of sensitivity to mean and variance in mPFC can, firstly, be replicated using an alternative neural measure provided by scalp electroencephalography (EEG) and, secondly, whether this timing of positive/negative relationship is suitable for a role in decision formation or feedback-related control. We report that this relationship can be observed at central midline electrodes, consistent with a mPFC source, and emerges on a timescale consistent with a role in decision formation, rather than suggesting an involvement in learning r updating. Furthermore, neural activity was again best described by the LPR model, providing further validation for the proposition of the integration of probabilistic information.

Chapter 5 focuses in more detail on the representation of summary variance. Using a priming version of the multi-element averaging task, we show that response times are facilitated when the variability of a target array is matched by a preceding task-irrelevant prime array. This priming by variability emerges even with a very short prime-target interval (200ms). Further experiments demonstrate the selectivity of the adaptation to variability; no parallel adaptation to the level of the mean is observed. Several features of the data suggest that variance priming operates at the level of perceptual representation, and is somewhat subject to the spatial distribution of attention. Together, these results support the conclusion that the visual system is highly sensitive to the distribution of environmental input and that sensitivity to the input statistics can guide short-term perceptual adaptation with resultant improvements for categorical judgement.

Chapter 6 aims to identity the neural correlates of the behavioural priming by variability described in Chapter 5. Similar behavioural adaptation to demanding stimuli has previously been associated with a network of brain regions, centred on mPFC and dorsolateral prefrontal regions. However, several features of our behavioural data suggest that priming by variance may operate at the level of perceptual representations. We therefore use fMRI to look for brain regions that show a repetition suppression effect for trials with consistent prime-target variability. We report that several brain regions show a pattern of BOLD suppression that exactly mirrors the behavioural interaction, including visual cortex (ventral temporal region) and lateral prefrontal and premotor cortices. Although no causal

relationship can be demonstrated, these results show the importance of sensory cortices for behavioural adaptation and are consistent with a multi-level view of behavioural and cognitive control.

Chapter 7 aims to provide a summary of the experimental work presented in this thesis and draws together several consistent themes. Firstly, this work demonstrates that the performance of human observers in categorical judgements is influenced by two sources of perceptual uncertainty: the mean distance to category boundary of the visual input and the variability of this input. Secondly, support is provided for the neural implementation of a model of perceptual choice in which observers make their decisions on the basis of probabilistic representations. Possible neural coding schemes are discussed that could implement the transformation to probabilistic representation. Finally, the behavioural and neural adaptation to variance demonstrates the importance and informativeness of information at the summary level and suggests that the associated behavioural limitations may arise not due to hard-coded limitations of the decision process but from a mismatch between the statistics of the input and the current distribution of neural resources.

Chapter 1: Categorical Perception and Visual Organisation

This thesis aims address the nature of the representations underlying perceptual categorisation in human observers, and how they are transformed throughout the visual hierarchy. We will address this question using a variety of methods, including behavioural measures, computational modelling, fMRI and EEG. Primarily, we will propose that independent behavioural limitations can arise from dissociable sources of uncertainty acting on these representations. In this first chapter, we provide an overview of the principles underlying perceptual organisation, how they compare to theoretical models of optimal perceptual choice and how they can implemented in a biologically plausible manner to support categorical choice.

1.1. Hierarchical organisation is a key principle of the visual system.

Since the first steps in understanding cortical visual representations (Hubel, 1959; Hubel & Wiesel, 1959), it has been clear that the visual system is hierarchically organised. Thus, the first cortical stop in the visual pathway represents the most coarse and basic of visual information, with small receptive fields most likely suited for the detection of local contrast (Hubel & Wiesel, 1962). Divergent outflow from early visual cortex then becomes increasingly specialised and integrated with other modalities or functions. Although several divergent hierarchies can be described in the cortical visual system (Goodale & Milner, 1992; Kato, Takaura, Ikeda, Yoshida, & Isa, 2011; Mishkin & Ungerleider, 1982), following any forward pathway will result in increasingly more complex levels of neuronal selectivity.

Although hierarchical representation is feature of all sensory systems (Duffy & Burchfiel, 1971; Iwamura, 1998; Rauschecker, 1998; Rauschecker & Tian, 2000), it has been most clearly demonstrated in the visual system through descriptions of the receptive field properties of neurons in different cortical regions. Neuronal receptive fields describe the area of external space that needs to be stimulated in order to provoke a response in any given neuron. Even just measuring the size of the receptive fields of different neurons shows a clear hierarchy; neurons in the earliest visual cortices have the smallest RFs, while neurons at more distant cortical locations have larger receptive fields (J. F. Baker,

Petersen, Newsome, & Allman, 1981), sometimes covering entire hemifields. There is, however, a necessary trade-off between specificity and generality that occurs within a single neurons. If a single neuron's RF covers a wide region of space, then that neuron will be insensitive to the exact location of a visual object within that wide region. A neuron with a very small RF will, conversely, provide a very specific signal about the location of a visual object, but will be less influenced by stimuli in other portions of the visual field (although note that these cells are not necessarily "blind" to other locations - see (Allman, Miezin, & McGuinness, 1985), with significant functional implications (Rao & Ballard, 1999)).

One of the advantages of hierarchical representation is that is can, within a few steps, produce quite complex response profiles from combining simple inputs in particular configurations. One clear example of this can be found at the first cortical synapses in V1. Neurons from the lateral geniculate nucleus have small, centre-surround receptive fields, ideally suited for detecting localised points of light (Hubel & Wiesel, 1962). These neurons are pooled in V1, resulting in so-called "simple cells", for which the ideal driving stimulus is an oriented strip of light. This apparently minor change in RF properties now allows these neurons to detect the high contrast contours and edges present in the visual field - which can be seen as a very rough sketch of spatial divisions akin to a statistical decomposition of an image (Bell & Sejnowski, 1997). Simple cells are then themselves pooled to create a new layer of complexity ("complex cells"). These cells are also orientation selective but can show a degree of spatial invariance, that is, by organising the input from several simple cells into a particular configuration (Fig 1.1), they become sensitive to particular contours at any location within their receptive field. This level of selectivity has been argued to be an important early stage for motion detection. Therefore, incremental pooling of neurons with similar response profiles can rapidly produce neurons with quite complex response profiles for analysing the visual input at a level more commensurate with guiding behavioural responses (Kourtzi & Connor, 2011; Tanaka, 1996).

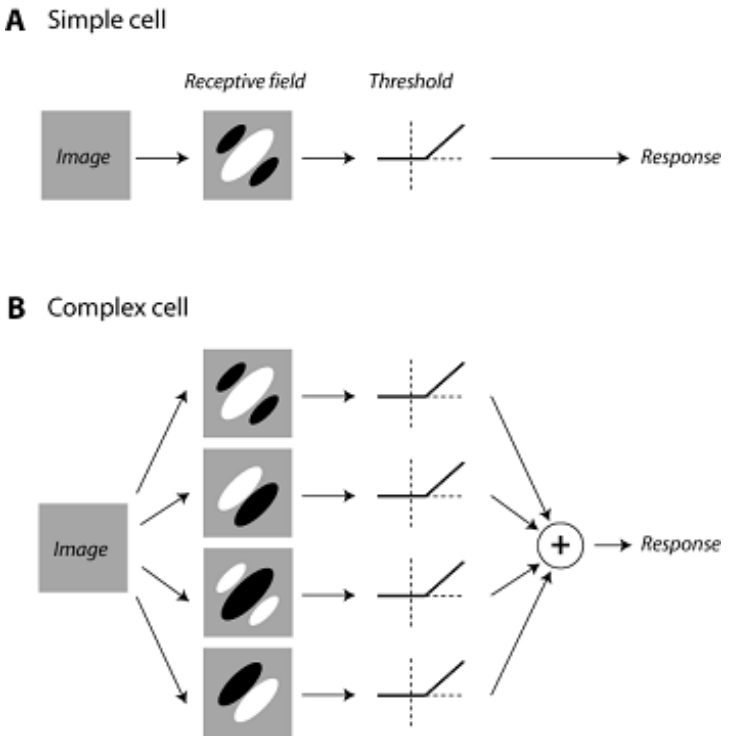


***Fig 1.1**. Model of simple and complex cells, after Movshon and colleagues (e.g. (Movshon, Thompson, & Tolhurst, 1978), showing the simple organisation of a simple cell receptive field (A) and the forms of RF structure contributing to complex cell RF properties (B).*

Figure from (Carandini, 2006).

This intricate system of hierarchical representation can seem at odds with the often simplified manner in which we interact with the visual world. Although these simplifications discard the kind of resolution V1 neurons have, they often bring a multitude of advantages. Perhaps the most ubiquitous example is perceptual categorisation - the grouping of physically distinct sensory input into discrete families of related items. Categorical perception is a highly efficient form of representation, since it allows for ready transfer of knowledge from one category member to another (i.e. generalisation) without having any experience with a new item. It also provides our experience of the world with some stability - small deviations or differences can be ignored (Murphy, 2002; Smith & Medin, 1981). It can also produce simple rules for the mapping of discrete behavioural

responses to a continuous sensory representation. This latter function is perhaps the most fundamental role of sensory systems; to sort or organise sensory information into the most convenient format for interaction with the current and future environment.

Investigations of perceptual categorisation thus often disclose the nature of perceptual representation. Although the RF properties of neurons across the visual system have been extensively investigated, the greater challenge has been to link these properties with their influence on behaviour. This is an important step as it can determine whether the information present in any given neuron is actually used by the system in question. One highly influential line of research has aimed to compare the empirically observed sensitivity of observers with the theoretical sensitivity of neurons in brain regions hypothesised to carry potentially relevant visual information (see (Britten, Shadlen, Newsome, & Movshon, 1992; Haefner, Gerwinn, Macke, & Bethge, 2013; Nienborg & Cumming, 2009; A. J. Parker & Newsome, 1998)). However, human behaviour and neural activity are measured on very different scales, and therefore to make meaningful comparisons between these two levels it is necessary to use equivalent measure of sensitivity.

1.2. Theoretical models of perceptual categorisation - optimal observers and mechanistic implementations.

Several potential approaches can be used to measure the sensitivity of an observer to sensory events. Perhaps the most straightforward is to take a performance measure, such as accuracy or response time. If an observer correctly identifies a sensory event at 100% accuracy, then it would not be unreasonable to conclude that the observer is sensitive to that event. However, simple measures such as accuracy can be biased. If an observer is asked to determine whether a display contains a target stimulus T (detection task), they could guarantee to accurately identify all T events by responding T on every trial. Even though accuracy in this case would be high, the observer demonstrated no ability to actually distinguish between cases of T and no-T.

To address this limitation of some performance measure, theoretical approaches have aimed to describe the general solution for such problems. The goal of these models is to describe the process (and potentially the mechanism) by which the theoretically maximal performance could be achieved, given the available information. In the case of a simple detection task, the task of the observer is to decide whether a particular signal is present or not. The process of detection is complicated by the presence of noise, both in our environment (e.g. an obscured view or low light levels) and additionally from even the earliest stages our internal processing systems ((Barlow, 1956; Barlow, Levick, & Yoon, 1971; Lamb, 1980)). Therefore, any given observation could reflect a true signal embedded in noise, or just the noise itself.

Signal Detection Theory (Green & Swets, 1966; Tanner & Swets, 1954) (SDT, Fig 1.2) described the optimal solution for solving this problem. SDT compares the likelihood that the current observation is generated by random noise fluctuations or by the presence of a real signal. It assumes that the observer knows the distribution of likelihoods across the range of potential input values for the cases of noise alone and signal + noise. There will be points in this sensory space (where 0 represents a total absence of observation) at which an observation is ambiguous; it could be due to noise or to a weak signal. Intuitively, this problem is best solved by using a criterion value. Any input that exceeds this value can be classified as signal and therefore anything below the threshold is noise. One key contribution of SDT is that it provides a method for placing the optimal boundary, which maximises the discrimination sensitivity. Sensitivity in SDT is specifically defined at the the z-transformed difference between the rate of correct detections vs false alarms (when the report was positive but no signal was actually present). As an alternative performance measure, it is therefore able to provide an estimate of sensitivity that is free from any response bias. SDT also provides an explanation for why some perceptual discriminations are more difficult than others. When the level of noise is greatest, the overlap between the portion of sensory space associated with the likelihood of signal + noise or just noise will increase. Therefore, there is a much higher chance for incorrect classification.

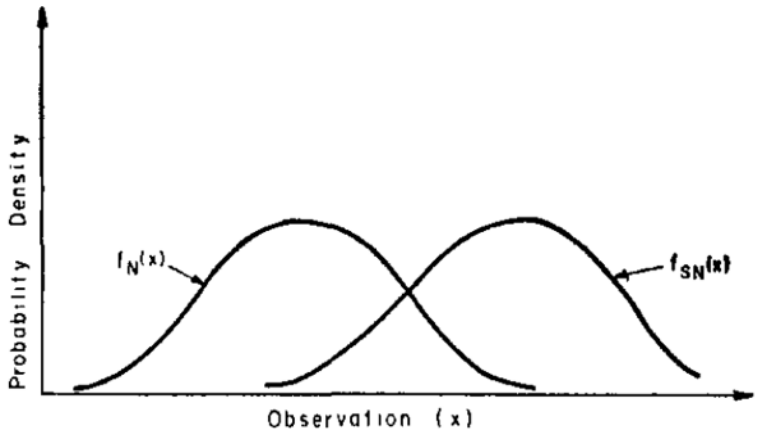


Fig 1.2. Signal Detection Theory*. From (Swets, Tanner, & Birdsall, 1961). SDT provides an optimal method for inferring the most likely category membership for single observation. Each point in the sensory space (x-axis) can be associated with a particular likelihood under distribution N or distribution SN. A criterion value can be imposed to minimise the errors (both false alarms and misses), which in this case would fall at the point of intersection between the two likelihood distributions. Response is determined by a comparison of the likelihood of the observation with this criterial value. This decision can be made more difficult (i.e. increased number of errors) either by increasing the variability of the likelihood distributions, or by bringing the mean points closer together.*

Although initially applied to the kinds of detection problem described here (Tanner & Swets, 1954), SDT can also readily be applied to categorisation problems. Instead of distinguishing cases of noise from signal + noise, the problem can become distinguishing Category A from Category B. The sensory input can be any particular feature space (e.g. colour, size, shape, motion direction etc), and the location of the criterion value will form a category boundary. Again, SDT shows us that the difficulty (i.e. likelihood of an error) of a categorical judgement will be determined by degree to which the likelihood distributions overlap. However, it is important to note that the degree of overlap can be determined by at least two independent manipulations. Firstly, the mean of the likelihood distributions can be brought closer together. Secondly, the variability of the distributions can be

increased. Although these manipulations are orthogonal, they will both result in decreased categorisation accuracy.

The appeal of SDT for those investigating perceptual categorisation is therefore that it outlines the optimal method for classification when there is any degree of uncertainty in the current observation. Since noise is a ubiquitous feature of all sensory environments, SDT can provide a theoretical framework for human perceptual categorisation. It has also been applied to obtain a measure of sensitivity from behavioural and neural data, and, critically, enables comparison between these two levels.

Parker & Newsome (A. J. Parker & Newsome, 1998) have argued strongly in favour of using these measures to make neural and behavioural comparisons to investigate which levels of perceptual representation are most closely linked with driving behaviour and perceptual experience. Although the general approach was attempted by many groups across several sensory modalities (Barlow et al., 1971; Mountcastle, LaMotte, & Carli, 1972; Siebert, 1965; Talbot, Darian-Smith, Kornhuber, & Mountcastle, 1968), perhaps the most most influential recent work has come from an extended series of experiments using motion detection as a model system in awake, behaving monkeys. Neural recordings were made from regions MT and MST (regions known to be crucial for perception of motion, (Dubner & Zeki, 1971; Newsome & Pare, 1988) whilst non-human primates (NHPs) performed a motion categorisation task. The pairing of brain region and stimulus is crucial - a neuron can only reasonably be expected to contribute to a task if, firstly, it is shown to be sensitive to the relevant feature dimension and secondly, the receptive field of that neuron is large enough to capture the required information (A. Parker & Hawken, 1985). This task requires the classification of the array on the basis of the average direction of motion of a cloud of moving dots. To manipulate the difficulty of the task, the percentage of dots moving in the same direction (motion coherence) can be varied. When the motion coherence is high (more dots moving in the same direction), performance on this task is both more accurate and faster than than when motion coherence is low.

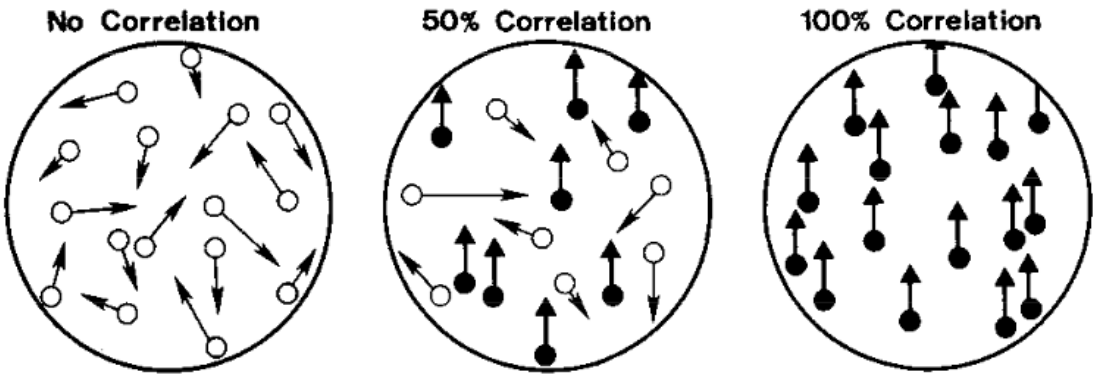


Figure 1.3. The random dot stereogram. *From (Britten et al., 1992). This task requires an observer to discriminate the overall direction of motion from a cloud of moving dots. To increase the difficulty of the task, the motion correlation (coherence) can be varied, from no coherent motion ("no correlation", left) to completely coherent motion ("100% correlation", right).*

Britten et al (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996) recorded from macaque MT whilst the monkeys performed the random dot motion discrimination task (RDM task, Fig 1.3). Recordings were made from neurons selected for their directional selectivity and the direction of motion to be discriminated spanned the full 360 degrees (although the discrimination was always between two exactly opposing directions). Motion coherence could range between 0% and 100%, the former case representing a completely ambiguous stimulus, where any evidence in favour of motion in one direction was balanced by motion in the opposite direction. As expected, the sensitivity of the monkeys was greatest for trials with the highest level of motion coherence. A parallel analysis of the neural data also revealed that the pooled activity across a population of neurons showed a near-identical sensitivity profile. The "neurometric" function displayed the the same sensitivity (slope) and any bias as the observer. This therefore suggests that these neurons provided information for the behavioural response.

More direct, causal evidence for the importance of these neurons in making the choice has been obtained by applying artificial stimulation to selected neurons (Salzman, Britten, & Newsome, 1990; Salzman, Murasugi, Britten, & Newsome, 1992). Applying stimulation to

“left” neurons was, for example, more likely to result in the monkey making the corresponding behavioural response. Therefore, this brain region is directly implicated in forming at least some part of the necessary neural representation of the category decision relevant information. This process is not limited to the visual pathways either, since a large body of work has revealed an parallel system for discrimination of vibrotactile frequency in the somatosensory system (Romo & Salinas, 2003).

One feature of note in the Britten et al data was that the response of neurons was unaffected by whether of not the monkey was actively performing the task, or just observing the stimulus array passively. This might suggest that MT neurons are not directly responsible for the categorical judgement - there might be an additional stage of processing in which the visual information is used to guide a response. Furthermore, many neurons were observed to have a greater sensitivity than the observer - see also (Vallbo AB, 1976) - perhaps suggesting that there is an additional stage of pooling that can introduce noise or somehow corrupt the perceptual signal.

Visual areas beyond MT have been investigated for the potential to represent information sufficient for perceptual categorisation. Since perceptual categories are often defined by object identity, many have studied the high level visual regions in ventral temporal cortex. Regions here fall into the ventral visual stream, a specialised visual hierarchy for the identification of visual object (Goodale & Milner, 1992; Ungerleider, 1982). Moreover, many regions contain neurons that show selective responses for particular stimulus classes. For example, neurons in superior temporal cortices inferotemporal cortex (IT) respond to 3D shapes, images of objects and faces in a selective fashion (Foldiak, Xiao, Keysers, Edwards, & Perrett, 2004; Freedman, Riesenhuber, Poggio, & Miller, 2003; Hung, Kreiman, Poggio, & DiCarlo, 2005; Kayaert, Biederman, & Vogels, 2005; Logothetis, Pauls, & Poggio, 1995; Logothetis & Sheinberg, 1996; Tanaka, 1996). Object responses are often view invariant, suggesting these neurons really code for the identity of the object. Therefore, for categorisation based on visual identity, anterior temporal cortex contains object level representations that could likely contribute to the categorisation process.

Kiani et al (Kiani, Esteky, Mirpour, & Tanaka, 2007) tested neurons in IT cortex with a broad set of stimuli to investigate whether these neurons contain representations useful for making categorical judgements. They measured the similarity of the neuronal responses to different images whilst the monkey performed a delayed match to sample task (based on the identity, not the category, of the image). Category membership was never necessary knowledge for accurate performance of this task so the representational structure in IT cortex was not cued by explicit task requirements. Using several methods to quantify the similarity of the population response across all pairs of stimuli, Kiani et al showed that images falling into the same "intuitive category" (e.g. animate vs inanimate objects) elicited more similar patterns than images from different categories. The overall structure also seemed to preserve some elements of a hierarchical representation – for example, more similar activity patterns were found for different images of hands than for different faces, although these groups were more similar to each other than to sub-categories of inanimate objects. Studies in human participants (Ewbank, Schluppeck, & Andrews, 2005; Haxby et al., 2001; O'Toole, Jiang, Abdi, & Haxby, 2005; Reddy & Kanwisher, 2006) using multivoxel analysis of fMRI data have similarly reported that the information in visual temporal regions is sufficient to support categorical distinctions. Kriegeskorte et al (Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008) compared the structure of these representations in both humans and monkeys, finding very similar categorical groupings of response profiles within inferotemporal cortex. Together, therefore, this has suggested that representations in IT cortex could support categorical judgements.

However, the contribution of this representational structure to human categorisation is unclear. Firstly, the results are difficult to interpret since the behavioural tasks (where present) do not necessarily require the observer to actually use these categorical distinctions. Therefore, it is not known whether these "categories" are just measures of perceptual similarity. A true categorical signal might be expected to show a representation that could be orthogonal to perceptual similarity, depending on the current task demands. However, fMRI studies have also shown that object selective cortices are activated when

human participants are asked to make categorical judgements about images of real-world objects (Martin, Wiggs, Ungerleider, & Haxby, 1996).

Some studies have proposed that in fact, there is more to the representational capacity of visual temporal cortex than the high level representation of individual object identity. For example, the selectivity of these neurons is modulated following paired-associate learning, such that the responses to each member of the pair become more similar following training (Erickson & Desimone, 1999; Messinger, Squire, Zola, & Albright, 2001; Sakai & Miyashita, 1991). This kind of phenomenon is perhaps also reflected in so-called statistical learning, a form of learning in which independent visual objects can become associated by covariance (Schapiro, Gregory, Landau, McCloskey, & Turk-Browne, 2014; Schapiro, Kustner, & Turk-Browne, 2012; Turk-Browne & Scholl, 2009; Turk-Browne, Scholl, Chun, & Johnson, 2009). Although this is some distance from a categorical representation, this ability of visual neurons to represent not only the current input but also its context is one mechanism by which perceptually dissimilar inputs can be grouped by another feature (in this case, their temporal location) that contribute to their identity. Furthermore, some neurons show learning effects for combinations of features, suggesting again that these high-level visual neurons come to reflect the functional groupings of features in the world (C. I. Baker, Behrmann, & Olson, 2002). However, this kind of modulation takes place in response to actual changes in the structure of visual information and does not reflect the kind of arbitrary boundaries that can be imposed by artificial categorical structures.

Therefore, it appears that the population of cell responses in inferior temporal cortex can provide information about the similarities of objects to other examples and perhaps can be modulated to reflect associations in time. Whilst categorical information could be derived from IT regions, it could be that an additional processing stage is required to translate this information into the required categorical groupings.

Together, these findings suggest that high level visual cortices represent perceptual information in a format that can be used to guide categorical judgements. However, as discussed, there are several reasons to suggest that a further stage of processing is needed before the information in these regions can be used to guide behaviour. For example, the

response of visual regions is largely consistent over presentations and any change occurs as a result of repeated change in the structure of the world. Functional categories, however, can be redefined without any corresponding change in the structure of the environment. Furthermore, it is unclear how information can be combined across regions when a particular item is to be judged on either a combination of features from different stimulus dimensions, or is a complex object that contains variability on a single dimension. In both cases, visual cortices might represent individual features or components, but it is unclear how this information is integrated.

A framework for understanding these distinction has come from the development of theoretical models of the decision process. All models in this lineage are based fundamentally on the principles of SDT. One advance, however, was the description of a statistical procedure for determining the likelihood of a particular hypothesis (e.g. Is the apple red?) on the basis of several, sequential sources of information. The Sequential Probability Ratio Test (SPRT (Wald & Wolfowitz, 1948)) calculates the cumulative likelihood based on all observed evidence, and will continue to accept new information until a threshold probability has been reached. This procedure will therefore find the most likely response option in the shortest number of steps. This approach inspired a new generation of models that were explicitly built to model perceptual decision processes (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). The crucial advantage of these later models is that they were able to predict distributions of response times based on any given configuration of inputs. These behavioural predictions can then be directly compared to empirically observed data. Although this class of models has become quite diverse in the precise mechanism suggested, they will be considered to broadly implement the same approach.

Perhaps the most widely adopted model in this class is the drift diffusion model (Ratcliff, 1978; Ratcliff & McKoon, 2008). The DDM represents the decision process as the diffusion of a particle towards one of two potential boundaries. Each boundary represents a behavioural choice (e.g. left vs right motion, see Fig 1.4). The observer is assumed to take samples of the current perceptual input over time, and the particle will move in the direction dictated by the evidence provided by each sample. Over time, the movement of

the particle will follow the dominant class of input. We can therefore make a distinction between the information provided by the current sample, and the overall movement of the particle representing the evolving decision process. The quantity reflected by the particle has been termed the "decision variable" (DV), since it is the hypothetical quantity upon which the decision will be made.

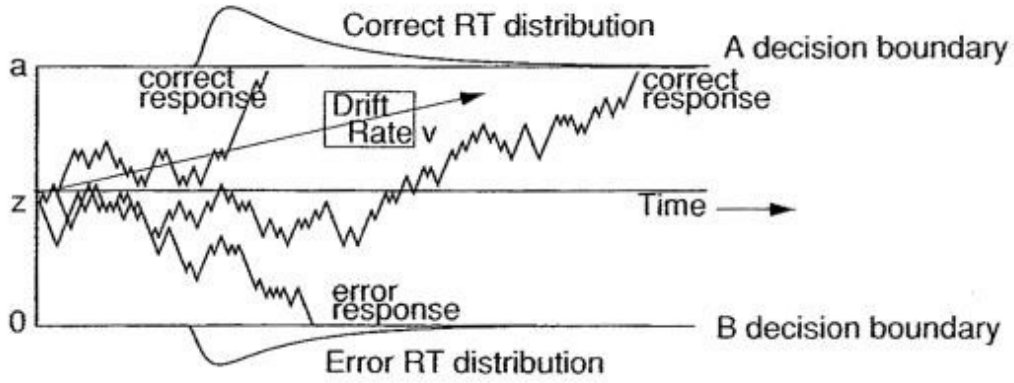


Fig 1.4. The drift diffusion model*. Figure from (Ratcliff & McKoon, 2008). The DDM proposes that the decision process can be modelled as the motion of a particle (starting at point z), moving towards one of two potential decision boundaries (A or B). Lines indicate path of motion. When the input to the model is noisy, the rate of accumulation (drift rate) will be slower, and the overall path to boundary will be longer. The predicted RT distributions for correct and error trials are shown overlaid at each decision boundary.*

Other models in this class have implemented the decision process with alternative structures (although formally many of these models reduce to an identical form (Bogacz et al., 2006)). For example, so-called race models (Vickers, 1970) integrate evidence for each response option as two independent, but competing, tallies of information. In this case, the DV can be characterised as the difference across two different pools of information. Other models have been motivated more by using biologically plausible structures, such as pools of mutually inhibiting neurons (Usher & McClelland, 2001; Wang, 2002).

Nevertheless, these models are able to capture many features of empirically observed behaviour. Primarily, they predict that response times should increase as the quality (often the signal to noise ratio) of the perceptual evidence decreases. For the DDM, this is because a noisy input will be more likely to be reflected in contradictory evidence on sequential samples. This means that it will take longer for the DV to reach one of the boundary levels of required evidence. Although there are several demonstrations of violations of this principle (Frank, Samanta, Moustafa, & Sherman, 2007), such as evidence for “fast errors”, the DV is most commonly seen to rise more slowly when the uncertainty associated with the perceptual evidence is low.

We can therefore use the framework provided by these models to specify the contribution of a particular neural representation to the categorisation process. These models predict that there should be signals reflecting momentary evidence that can be distinguished from an evolving DV. It is the DV that therefore reflects representation at a more categorical level. However, although this theoretical framework provides a neat conceptual distinction between momentary evidence and an accumulation of this evidence in the DV, it is unclear whether the brain uses a similar mechanism. To investigate whether the brain does implement a form of accumulation-to-bound, many studies have attempted to find a neural signal that reflect the expected properties of a DV.

1.3. Comparison of Theoretical Model Predictions and Functioning of Neural Systems.

A neural implementation of an accumulation to bound model would need to display certain key features. This signal would also be expected to increase in amplitude over time, and the rate of this increase should be dependent on the quality (e.g signal to noise ratio) of the total evidence presented, rather than momentary fluctuations in perceptual information. It might be expected that here should also be evidence for a kind of common threshold across decisions.

A series of experiments by Shadlen and colleagues has used the primate motion discrimination pathway as a model system to investigate these principles. Building on the previously described work in the primate visual system on the representation of motion,

this group have tried to identify cortical circuits that reflect different stages of the decision process. Their hypothesis was that the types of visual region previously studied (e.g. MT) represent the moment-to-moment perceptual evidence. Although these neurons theoretically contain the information required to support the observed behavioural sensitivity, it was argued that a separate stage is needed to translate this sensory representation into a discrete response of the type required by the motor system (Gold & Shadlen, 2007). One candidate region for performing this intermediate stage is the parietal cortex. Anatomically located between sensory-specific and motor cortices, it has long been implicated in sensory-motor transformation (Andersen & Buneo, 2002) and attentional selection (Colby & Goldberg, 1999), arguable highly relevant cognitive functions for categorisation.

When Shadlen and Newsome (Shadlen & Newsome, 2001) recorded from neurons in lateral intraparietal cortex (LIP, chosen specifically for its role in visual guided eye movements) as monkeys performed the RDM task, they observed that the overall level of activity increased throughout the trial, as opposed to the steady sensory response to the RDM stimulus in MT. The spatial location of the receptive field of each neuron could be classified as either being congruent or incongruent with the direction of motion in the stimulus. For example, a neuron with a receptive field on the left of the screen would be congruent on trials with overall leftward motion, but incongruent on trials with overall rightward motion. The responses of individual neurons increased on trials with motion that was congruent with the location of their receptive field. Therefore, these neurons were responding not to the physical stimulus but in a manner that signalled the eventual response. At the population level, this selectivity became clearer. Over the course of a single trial, neurons with RFs at the correct location for response slowly ramped up their activity until they reached a threshold level, after which the response was initiated (Fig 1.5). This threshold was determined by fixing the neural activity traces to the response, rather than the stimulus, where it was found that each trace peaked at an identical level of firing rate at a similar time before response initiation (although see (Kira, Yang, & Shadlen, 2015)). Neurons with RFs at the incongruent location showed a slight reduction in activity over time.

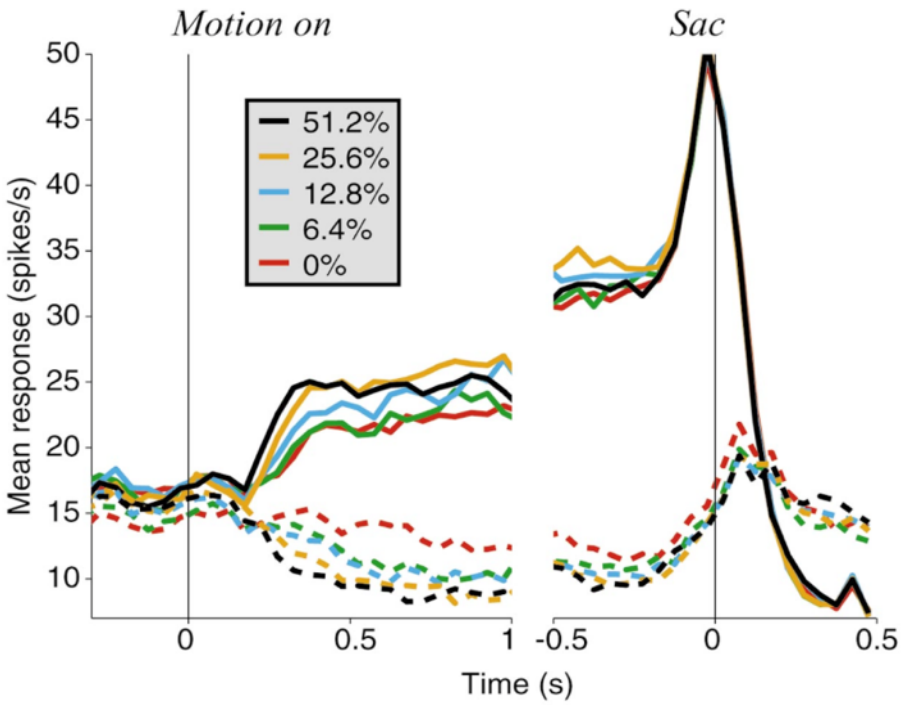


Fig 1.5. Response of neurons in LIP during the response phase of a RDM task. *From (Shadlen & Newsome, 2001). Left: shows average neural firing rate (y-axis) locked to the onset of a RDM stimulus. Coloured lines indicate different level of motion coherence. Solid lines indicate choices made into the response field of the neuron, dashed lines indicate response on trials for non-preferred response field. Right: Lines as previously. Mean response of neurons in this case is locked to the moment of response (a saccade).*

Perhaps the most striking feature of this data, however, was the relationship between the rate of increase of congruent location activity and the strength of motion in the stimulus. Motion strength was manipulated so that many of the dots were moving in the same direction (coherent motion) or there was an increasing number of dots moving in alternative directions to the dominant direction (incoherent motion). When the population responses were sorted according to motion coherence, the build-up rate in response congruent neurons was significantly faster for higher levels of motion coherence. Since it appeared a threshold level of activity (i.e. all neurons reached the same level of activation before response) was required before a response was initiated, this meant that on trials with lower motion coherence, the population response took longer to reach this threshold

level. Since these trials are also those for which the slowest and least accurate behavioural response is observed, Shadlen and Newsome concluded that this build-up in activity is the likely determinant of the final categorical response.

Many subsequent studies have extended these findings, providing causal evidence for the role of LIP in representing the DV through direct electrical stimulation (Hanks, Ditterich, & Shadlen, 2006) or by looking at disruptions in the DV introduced by controlled perturbations in the sensory evidence (Huk & Shadlen, 2005). Additionally, this process appears to be sensitive to non-perceptual features of the task, such as the value of the perceptual categories (Platt & Glimcher, 1999) and is independent of the motor plan required to carry out the response (Bennur & Gold, 2011; Rishel, Huang, & Freedman, 2013). Together, this work suggests that there is a distinct neural circuit which supports the pooling of perceptual information over time to form a discrete categorical output. LIP neurons also appear to be sensitive to multimodal information (Cohen, 2009), and more recent analytical techniques have demonstrated the range of information (both decision relevant and irrelevant) that can be decoded from the population response (Meister, Hennig, & Huk, 2013; Park, Meister, Huk, & Pillow, 2014). However, the “ramping-to-bound” response characteristic of neurons in LIP is not unique to this brain region. For example, this kind of pre-response activity has been observed in dorsolateral prefrontal cortex (Kim & Shadlen, 1999), the superior colliculus (Horwitz, Batista, & Newsome, 2004) and even in the caudate nucleus (Ding & Gold, 2010). This response profile may therefore be a feature of neuronal computation throughout the associative brain.

Perhaps therefore, this response profile alone is not sufficient to determine the contribution of any given brain region to perceptual choice. One useful feature of categorisation is that the definition of a category can be flexible - the criteria for membership can updated or changed based on current demand or experience. If a brain region were representing information in a categorical sense, therefore, we would expect to see evidence for this flexibility. To address the question of representational flexibility, therefore, a useful stimulus set should dissociate between perceptual features or perceptual similarity and categorical identity.

1.4. Neural studies of categorisation have demonstrated the distributed nature of the representations necessary for categorisation.

Freedman et al (Freedman, Riesenhuber, Poggio, & Miller, 2001) created a stimulus set to test exactly this form of representational flexibility. Their stimulus set used six prototypical animals (three cats and three dogs) and further test stimuli generated by morphing between the six prototypes. These images therefore existed on something approximating a continuous scale, with equal "perceptual similarity" between neighbouring images. A category boundary could then be placed at any point along this scale. Freedman et al initially trained their two monkeys to categorise cats vs dogs in a delayed match to sample task, with the boundary placed at a central location. Note that this structure means that pairs of stimuli within the category can be more dissimilar than pairs of stimuli across the category boundary, providing the dissociation of perceptual similarity and category membership.

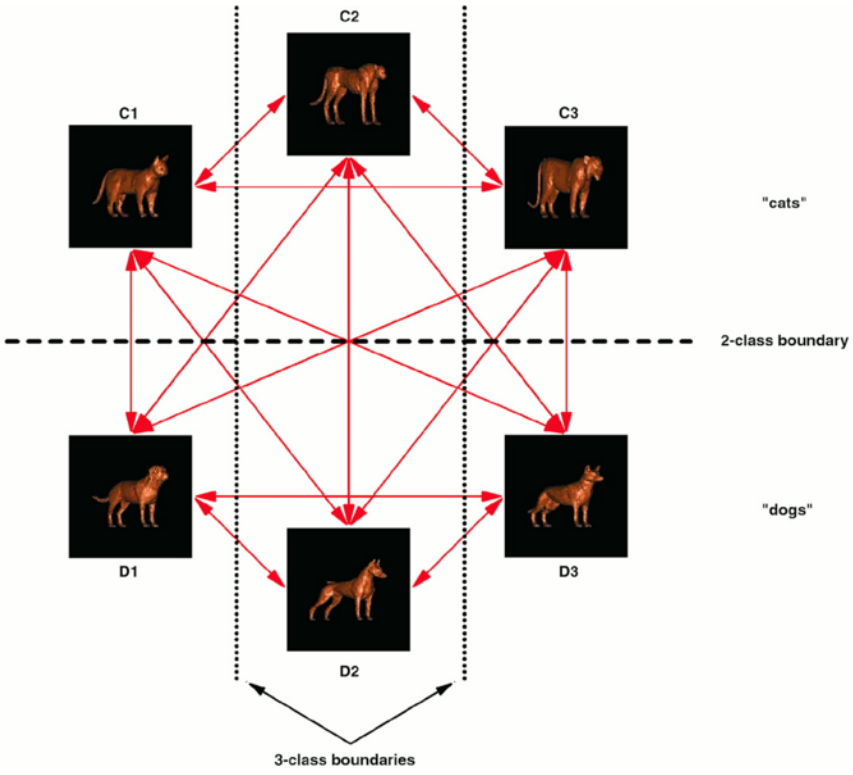


*Fig 1.6. **Continuously varying stimulus set for categorisation tasks**. Figure from (Freedman et al., 2001). Stimuli were generated by morphing prototypes to form a continuous set of test stimuli. An arbitrary category boundary (or boundaries) could be imposed at any point in this morphed space.*

Recording from lateral prefrontal neurons, Freedman et al showed that not only were there neurons that showed category selectivity (i.e. preferential responses to stimuli from one category) in both the stimulus and test periods, but these same neurons did so irrespective of perceptual similarity (Fig 1.7). This meant that category selective neurons showed similar firing rates for stimuli that came from the same category, despite perceptual dissimilarity, and significantly different firing rates for pairs of stimuli from different categories, despite some degree of perceptual similarity. Moreover, the same animals were later re-trained to divide the stimulus set into three categories and prefrontal neurons were observed to follow the same re-mapping of categorical definitions. Freedman et al concluded therefore that these patterns of firing show lateral prefrontal cortex to be a region capable of supporting true categorical representation.

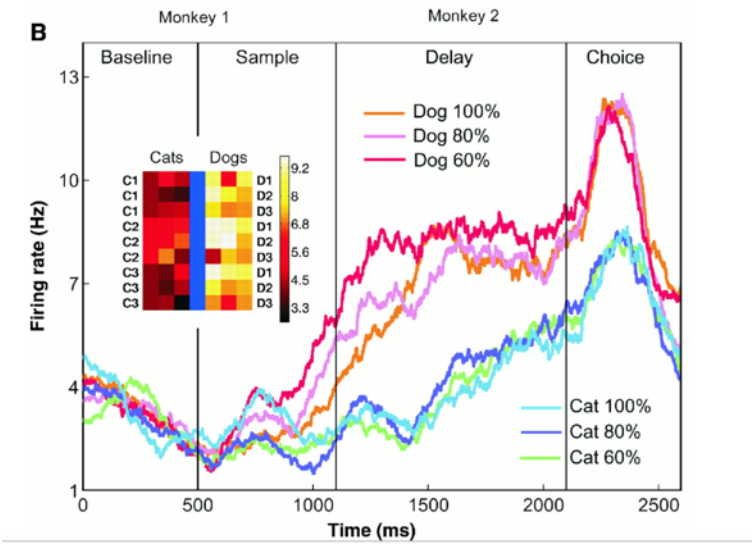


Fig 1.7. Results from (Freedman et al., 2001) showing categorical representation by prefrontal neurons. Lines indicate firing rate (y-axis) across time, locked to the onset of the trial. Colours indicate category condition (hot colours = dogs, cool colours = cats).

A later experiment (Freedman et al., 2003) used the same stimulus set in ventral temporal cortex and failed to find the same kind of categorical response. This would be consistent with the view outlined above, in which ventral temporal cortex supports representations of identity in terms of their perceptual features or feature conjunctions, but it is only associative cortices that support categorical representations. However, the behavioural performance of their monkeys in the original study shows little difference in categorisation performance for prefrontal neurons across the levels of morph used in their study. This perhaps suggests the monkeys were overtrained compared to what would be expected from relatively naive human observers. It would be expected that the most blended items would be the most difficult to categorise. For the highest morph level (60% cat/40% dog or vice versa), performance was at approximately 90%, suggesting little difficulty. It is therefore not clear whether these neurons are representing a behavioural rule or an estimate of category membership.

A similar approach was used to investigate the categorical responses of neurons in LIP (Freedman & Assad, 2006). In this case, monkeys were trained to categorise RDM stimuli with respect to a particular category boundary. As expected, LIP neurons demonstrated selectivity for a particular category. Importantly, after re-training with a new category boundary, the same neurons shifted their selectivity to reflect the new category structure. This was not the case with neurons in MT, which always faithfully encoded the actual motion direction, irrespective of the current category boundary. Therefore, neurons in parietal cortex demonstrate the kind of flexible representation that would be expected from the computation of categorical responses. Furthermore, comparison of temporal dynamics suggest that these categorical signals emerge early in parietal cortex, meaning they are unlikely to reflect feedback from prefrontal regions (Fitzgerald, Swaminathan, & Freedman, 2012). In contrast, sensory cortex represents information with respect to the structure of the external world, rather than the internal divisions of categorical structure.

Most of the work presented so far has been derived from single unit studies in non-human primates. A growing literature however has been using neuroimaging and electrophysiological recordings to investigate perceptual categorisation directly in the human brain. Although these methods lack the spatial resolution of single unit studies,

their global nature has demonstrated that categorisation relies on an interconnected set of brain regions, potentially working in parallel.

1.5. Human neuroimaging studies have found several neural correlates for decision variables, and have highlighted the interconnected cognitive processes underlying accurate perceptual choice.

Much like the single unit literature, neuroscientific studies of human perceptual categorisation have largely been concerned with the identification of regions that represent the theoretical decision variable. Again, similar to the single unit primate studies, there seems to be a broad dissociation between brain regions that represent this momentary sensory evidence and those that represent the ongoing DV, calculated from the integration of evidence over time (see (Heekeren, Marrett, & Ungerleider, 2008). For single unit studies, the key characteristics of a decision variable signal have been a rate of build up that corresponds to the integrated evidence, rather than the momentary demand, and that this rate should reach a particular value before a response is initiated (Gold & Shadlen, 2007). Translating these model-based predictions into predictions for how human neuroimaging methodologies should vary with decision difficulty has not been straightforward (O'Reilly, Jbabdi, & Behrens, 2012; O'Reilly & Mars, 2011; Serences & Saproo, 2012), although this model-based approach can potentially provide more specific hypothesis for often ambiguous neural signals.

For some of the first fMRI studies looking for a neural correlate of the DV, the debate about the expected direction of BOLD modulation by the DV became (and still is) a crucial issue. Heekeren et al (Heekeren, Marrett, Bandettini, & Ungerleider, 2004) used the principles of of the drift diffusion model to make predictions about the BOLD signal. They argue that a brain region reflecting the integration of sensory evidence show show the greatest BOLD signal on trials with high levels of sensory evidence (i.e. the least noise), since these trials are associated with the steepest rate of increase in the DV. This faster rate of increase should therefore be associated with greater metabolic activity, and therefore a greater BOLD signal. They report several regions that follow this prediction, including the dorsolateral prefrontal cortex. This region was argued to be a viable candidate for a

representing the final decision, since the signal here was also dependent on the difference in the available evidence for each category. Other work has argued for a role of human IPS in stimulus-response mapping, on the basis of a negative relationship between the amount of sensory evidence and BOLD signal (Tosoni, Galati, Romani, & Corbetta, 2008).

However, other theoretical work suggested that one foundational assumption behind the interpretation of these results – that of the negative relationship between BOLD and decision difficulty for regions representing a decision signal – was unsafe (Basten, Biele, Heekeren, & Fiebach, 2010; Heekeren et al., 2008; Ho, Brown, & Serences, 2009; Kayser, Buchsbaum, Erickson, & D'Esposito, 2010). In fact, it was proposed that this assumption should be inverted, meaning that a greater BOLD signal should be expected in these regions in response to increasing levels of decision difficult. This argument was made firstly on the basis of neural data showing that the overall firing rate during a trial for decision-sensitive neurons was greater under noisy (i.e. low evidence) conditions, largely due to the greater amount of time taken to reach the decision threshold, and secondly it was suggested that RTs should scale positively with BOLD signal.It might be expected that neural structures should be energy-efficient; easy tasks are by definition less demanding and therefore an efficient system might minimise the metabolic costs of these cases, reserving energy for more demanding scenarios. Several brain regions show this relationship with stimulus noise, including the types of parietal region so well studied in the primate literature (Heekeren, Marrett, Ruff, Bandettini, & Ungerleider, 2006).

However, this positive relationship between BOLD signal and uncertainty cannot be taken as a sufficient criterion for the identification of decision variable signals, since many cognitive factors may be expected to correlate with decision difficulty. For example, as decision difficulty increases, networks associated with cognitive control might be expected to become increasingly engaged. The aggregate neural activity indexed by the BOLD signal is one method of assessing brain activity during the decision process. MEG and EEG have also been used to investigate cortical representations of potential decision variables. Although these methods lack the spatial specificity of fMRI, they can provide insight into the temporal dynamics of the decision process.

Given that one feature of a decision variable-like signal should be that it grows over time in proportion to the quality of the evidence prior to response, electrophysiological methods have the potential to capture these kinds of relationships. Donner et al (Donner, Siegel, Fries, & Engel, 2009) used MEG to record activity whilst participants performed a motion detection task. Subjects reported the presence or absence of motion with different hands, so the authors were able to take advantage of lateralised motor signals corresponding to each response. In this task, motor lateralisation was observed as an increase of power in the gamma band and a decrease in power at lower frequencies. Importantly, this activity gradually built up in the pre-stimulus period, with greater degrees of lateralisation associated with a greater probability of choosing "yes" with that hand. Furthermore, this activity correlated with a motion integration signal arising from motion sensitive visual regions. This suggests that, for this task, the visual information is directly integrated to form a motor plan. De Lange et al (de Lange, Rahnev, Donner, & Lau, 2013) found that a similar lateralised motor preparation signal became apparent at a greater rate for higher levels of motion coherence, replicating the relationship between increase in decision variable and decision uncertainty seen so many times in single unit studies.

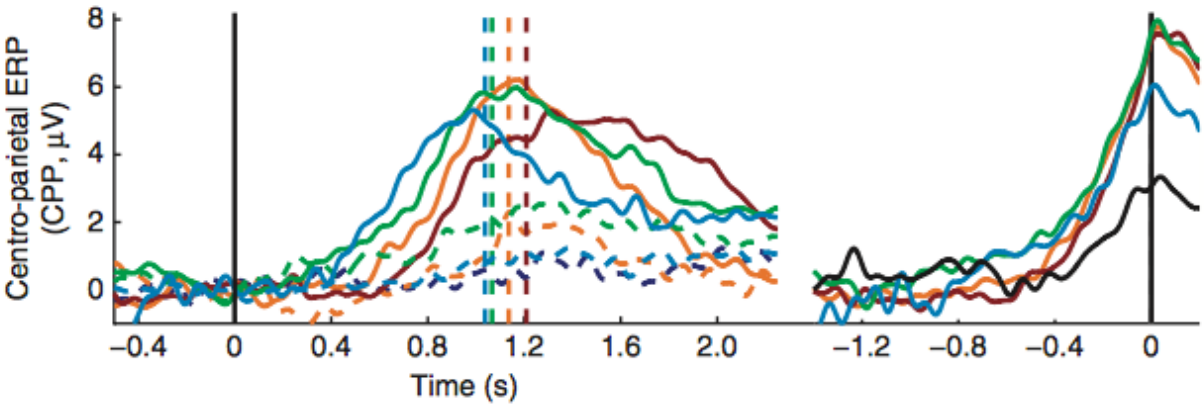


Fig 1.8. The centroparietal positivity (CPP). *Figure from (O'Connell, Dockree, & Kelly, 2012). Coloured lines indicate the amplitude of ERP amplitude over time, for different levels of stimulus-to-noise ratio. ERPs are plotted for the stimulus locked data (left) or response locked (right).*

A more recent set of EEG studies has attempted to isolate decision signals that are independent from the specific sensory or motor demands of the current task. O'Connell et al asked subjects to monitor a constant flickering stimulus for changes in contrast. The flickering of the stimulus meant that the neural signals associated with momentary changes in stimulus contrast were represented by changes in amplitude within a corresponding frequency band. Since motor signals can be identified by their lateralised distribution, O'Connell et al (O'Connell et al., 2012) reasoned that any remaining signal was not related to either motor preparation or the direct encoding of sensory information, and so was potentially a reflection of the decision process itself. They describe a central-parietal positivity (CPP, Fig 1.8), which was a slow build up in amplitude occurring at central/parietal electrodes prior to the behavioural response. This signal peaked at approximately the same amplitude before response, suggesting a form of threshold mechanism. Additionally, this signal was associated with false reporting of target detection, demonstrating the behavioural relevance of this signal even in the absence of the appropriate sensory evidence. A follow up study replicated many of these findings using a RDM task, although modified to mean that participants were continuously monitoring a stream of information, rather than presented with clearly distinct trials. With this paradigm, Kelly et al (Kelly & O'Connell, 2013) were able to demonstrate the consistency of the CPP across modalities and its association with the strength of perceptual evidence, suggesting that this signal potentially reflects the integration of perceptual information into a decision variable, independent of direct motor planning.

Human neuroimaging has therefore demonstrated that similar integrative neural mechanisms can account for perceptual categorisation performance across species. Although the exact nature and distribution of this integration appears to fluctuate with the specific task demands, describing categorisation as the end point in a process of evidence accumulation can account for many features of behaviour and neural signals. However, this integration of evidence is not the only process associated with forming a perceptual choice. In a review of neuroimaging work on human perceptual decision making, Heekeren et al (Heekeren et al., 2008) propose a set of serial processes that function together for effective decision making. These modules are not intended to be seen as completely

independent but nevertheless it is proposed that they rely on different brain areas and perform different functions. The demarcation of these modules is based on both the single unit and human imaging data described above. The first module is responsible for the accumulation and comparison of sensory evidence, the second is responsible for the detection of perceptual uncertainty (and can implement strategies to re-allocate resources appropriately), a further module represents the decision variable and finally, a fourth module is responsible for performance monitoring (responsible for the detection of error and a consequent need to a change in decision strategy).

Although the distinction between these modules is not necessarily concrete, this scheme highlights the multitude of contributing factors to the formation of a single perceptual choice. One often overlooked question is the representation of the sensory evidence itself. Although often assumed to be supported by neurons with relatively stable sensory preferences, there is a large body of evidence to suggest these preferences are constantly fluctuating to maximise the efficiency of representation. These adjustments can be seen at the level of sensory representation, but may also reflect the operation of systems intended to monitor ongoing performance for the purposes of error correction.

1.6. Cognitive theories of categorisation have argued for several different forms of categorical representation.

The work discussed so far has largely aimed to describe the neural mechanisms underlying the conversion of sensory input into clearly defined response categories. As discussed, however, it has proven difficult to identify the nature of the unique contribution of each identified brain region or set of regions. A literature in cognitive psychology, however, has aimed to identify exactly the nature of the representations that contribute to categorisation.

One of the most important advances was to recognise the graded nature of category membership. For example, although we might consider the category “cup” to be based on simple definitions of features (e.g. has a handle, holds liquids), there are some category members that would be considered more representative than others (e.g. teacup).

Furthermore, there would be some disagreement about whether some items (e.g. small soup cup) even belong in this category (Smith & Medin, 1981). Empirically, human participants are likely to endorse more “prototypical” items as being members of that category with a shorter response time than for other categories (Rips, Shoben, & Smith, 1973). Although these typicality effects could represent a calculation of the number of features an item possesses, it can be shown that humans will identity some objects as more representative than others of a particular category, even when all items share the same number of necessary features (Armstrong, Gleitman, & Gleitman, 1983; Lupyan, 2013).

These findings challenge earlier views, in which the properties of an item are compared to a stored list of necessary and sufficient features for that class. Although comparison must form a crucial stage of category determination, perhaps the more relevant question concerns the nature of the internal representation against which an item must be compared. Although this question has been largely ignored by those focusing on the neural mechanisms of perceptual choice, a large literature in cognitive science has attempted to determine the nature of categorical representation.

The outcome of several decades of research resulted in two suggested structures for graded categorical representation. The first suggests that the stored representation reflects the average of all previously experienced category members, termed a “prototype”. Experimental evidence has largely relies on tasks in which observers are asked to learn to categorise examples from a set of objects into two classes. All examples have been generated by distorting a prototype by varying degrees, although the participants never have direct experience with the prototype. Performance on a new set of items is then tested, with the typical finding that observers are able to accurately classify these new items, even though they have never seen them before (Posner & Keele, 1968). According to prototype theory, this is because participants were able to average across the distorted items to retrieve an accurate representation of the original prototype.

However, an alternative theory, exemplar theory, proposed that this task can be solved by storing each exemplar in memory and then calculating the distance between the new item

and all previously experience items. If the new item is closer to more items from Category A, then it will be classified as part of this category (Medin & Schaffer, 1978; Nosofsky, 1986). This form of representation has been associated with geometric projections of similarity, with some suggestion that the firing patterns of neurons in inferotemporal cortex can represent the similarity of object identities in exactly this format (Op de Beeck, Wagemans, & Vogels, 2001). Exemplar theories have several advantages over prototype theories, in that they are able to, for example, explain how exceptions are so often easily recalled or classified. Under prototype theory, exceptions are those that lie furthest from the extracted average and so should pose the most difficulty for the observer. Some have also argued that exemplar theories can produce prototype like-behaviour if the assumption is made that a judgement is reached by random sampling of stored exemplars, which are then pooled into a single representation.

Although many have tried to pit exemplar and prototype theories against each other in the same experiment, this has proven to be very difficult, with little agreement on the outcome (Blair & Homa, 2003; Medin & Schaffer, 1978). One problem is that these two theories often make identical behavioural predictions. Mack et al (Mack, Preston, & Love, 2013) used multivoxel analysis methods on fMRI data obtained whilst participants performed a categorisation task. This form of data analysis is suited for looking not at differences in the level of activation between conditions, but rather at the pattern of activation, making it highly applicable for the study of how information is represented by neural ensembles. To investigate the neural representation, Mack et al calculated a measure which they termed “representational match”, which measured the similarity between the current test item and the category representation of their models. This quantity reflects the implementation of an exemplar-based scheme of representation. They found that with this measure, the exemplar model better described neural activity in lateral occipital and posterior parietal cortex, compared to a prototype model. Therefore, it seems that the neural data could provide a dissociation not available at the level of behavioural performance.

Interestingly, however, performance on a transfer task in a prototype experiment can be improved when the training set included more highly distorted items (i.e. a greater

variability)(Posner & Keele, 1968). Although variability is usually assumed to increase noise, and therefore impair performance, in this case it might aid in helping observers to build a representation of not only the average stimulus, but also some idea of the full distribution of possible configurations associated with that category.

An alternative approach, developed from the principles of single detection theory, explicitly incorporates knowledge of the distributions of possible values associated with each category (Ashby & Gott, 1988). This decision bound theory suggests that if we can extract not only the mean but also the variance of the likelihood distribution over the relevant feature dimension for each category, then we can calculate the optimal decision boundary to maximise categorisation performance. As discussed earlier, performance by human observers in classification tasks is often well-explain by assuming they calculate the optimal decision boundary as explained by SDT, and can also account for accurate classification performance even when the decision boundary is non-linear.This type of representation could also presumably arise from the pooled integration of a sub-sample of exemplars, which would provide and estimate both of central tendency and the likely variability associated with that category.

This more statistical approach has broadened into a conception of categorical relations as a measure of the covariance among different stimulus properties. Connectionist models have successful learnt to categorise simple set of objects through sensitivity to the "coherent covariation" among the stimulus properties (McClelland & Rogers, 2003). This approach is able to replicate a remarkable number of features of human categorisation, in additional to demonstrating the same evolution of representation as observed in developmental studies with human infants. For example, the model displays the kind of graded representations observed in infant categorisation, and the development of these representations over model iterations follows the same broad-to-narrow trajectory as the development of category knowledge in early childhood (Rogers & McClelland, 2008).

Although these models are impressive in their explanatory scope, their success relies on repeated observations over time and the gradual accumulation of experience to appropriately estimate the relevant statistical relationships. One criticism of such models

is that they do not reflect the preferential treatment our cognitive structures give to some inferences; for example, children area able to infer causation in event relationship after very limited experience (Tenenbaum, Kemp, Griffiths, & Goodman, 2011). A more recent computational approach has argued that using Bayesian methods for understanding the inferences human participants make can overcome these limitations. For example, recent evidence has suggested that human infants show an ability to distinguish a new category on the basis of a single exemplar (Gerken, Dawson, Chatila, & Tenenbaum, 2015). These results can be explained by assuming that infants are able to track the likelihood of combinations of events, on the basis of an evolving model of the world. Although not incompatible with a mechanism that learns about the correlative structure, it suggests that categorical representation relies on a knowledge of the likelihood of particular feature associations. Highly unlikely combinations therefore provide an immediate cue that the current input reflects a new category.

1.7. Chapter Summary

Neural studies of classification and perceptual choice, in combination with clear computational models, have outlined a clear framework for the investigation of evolving categorical choice. A broad consensus has arisen which proposes category judgement arise following the accumulation of sensory evidence in favour of one or other relevant option. This mechanism has been observed across a number of paradigms and methodologies. Still, the nature of the representations on which these mechanisms operate is still unclear. The observation of similar neural response profiles across multiple brain regions has highlighted the lack of current understanding regarding the key component representations relevant for perceptual choice. The next chapter will discuss in some detail a body of work suggesting that the informative signals in sensory input are often to be found at the population, or summary, level.

Chapter 2: Summary Representation in Perceptual Systems

One question this thesis aims to investigate is the nature of the informative representations used in the processes underlying perceptual choice. Although models of perception are often based on the hierarchical processing of individual features, an alternative form of representation can arise at the summary level; incorporating information from multiple sources. This chapter will review the work that supports the idea of summary representations and their potential contribution to categorical judgements.

2.1. Perceptual averaging in human observers: the global level of perception as a statistical summary measure.

The ability of human observers to make an accurate response on the basis of an average percept has been known for some time. Observers are able to report the average direction of motion of a display of randomly moving dots (Williams & Sekuler, 1984), average speed (Watamaniuk & Duchon, 1992), as well as the average size (Chong & Treisman, 2003) of sets of visual objects (Fig 2.1). This ability even extends to more complex visual stimuli with high dimensional features, as shown by several experiments in which participants were asked to extract the average emotion from a series of faces (Haberman & Whitney, 2007, 2009). There is thus abundant evidence that the brain can generate a response on the basis of multiple, simultaneous inputs to the visual system.

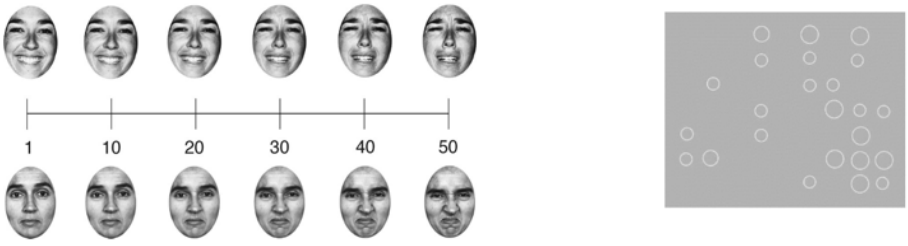


Fig 2.1. Examples of stimulus sets from which observes can accurately extract a summary percept. *From (Haberman & Whitney, 2007) for faces and (Chong & Treisman, 2003) for circles.*

Furthermore, this average, or summary, response seems to be sensitive to the uncertainty associated with the input. For example, the random dot motion task requires observers to respond to the average direction of motion of a set of rapidly moving dots. The speed and accuracy of response is dependent on the ratio of the dots moving coherently in the “true” direction against those moving in any directions. Thus, human observers are more accurate in identifying the dominant motion direction when 30% of dots are moving coherently than when the coherence is lower. There are therefore two kinds of information that can be extracted from a visual input - the identity of low-level features and summary measures that take all these features into account.

Perceptual averaging is perhaps reminiscent of a distinction often drawn between “global” and “local” levels of representation. Navon (Navon, 1977) argued for a similar concept of a distinction between local features (component parts) and global features (those that emerge from the integration of the components, Fig 2.2). Previously, Gestalt psychologists had argued that perceptual systems do not just respond to combinations of features but create a distinct representation of the integrated whole which represents the first stage of visual processing. This “whole-first” approach was proposed to explain why, for example, the overall spatial organisation between elements could guide interpretation of a group of objects. This view, however, was not commensurate with the later investigations into the neurophysiology of the visual system. Decades of research emphasised that visual processing began with a decomposition of the image into the most basic features and then proceeded through a hierarchy of increasingly more complex and abstracted levels of interpretation (Hochstein & Ahissar, 2002).

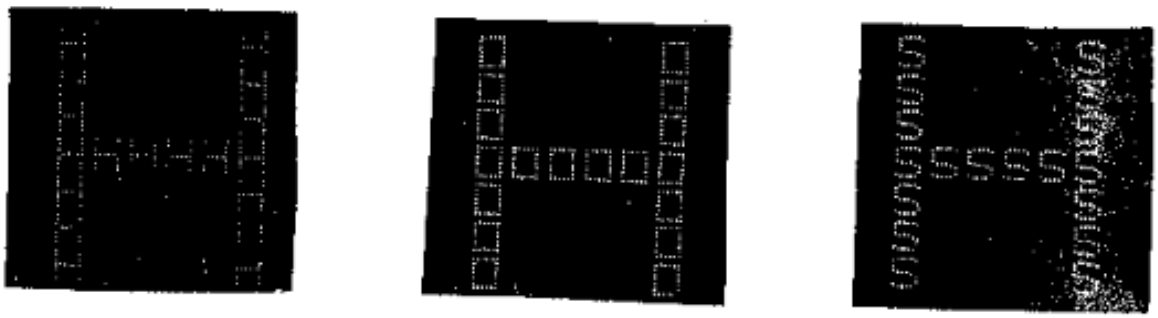


Fig 2.2. Hierarchical figures*. From Navon (1977). At the global level, all three images have the same identity (the letter "h"). Each letter is constructed from smaller letters at the local level, and these can be consistent (left), neutral (centre) or conflicting (right) with the global level.*

Navon's proposal of global and local processing aimed to somewhat reconcile these findings, although greater emphasis was placed on the global level as the primary, automatic stage of processing. For example, he created letter stimuli that were built from smaller letters (Fig 2.2). These two levels of the stimulus could therefore either be the same (large H from small Hs) or conflicting (large H from small Ls). When participants were asked to categorise an aurally presented letter sound (either congruent or incongruent with the visual information) whilst looking at such a visual stimulus, he found that the visual information only produced interference when it was the global level that differed from the auditory information. In a separate judgement task, participants were more likely to classify a pair of briefly presented letter stimuli as different if they differed on the global level rather than the local level. Although there are many demonstrations in the selective attention literature of the importance of local features, experiments such as Navon's highlight that the global information about a scene can take precedence over a more detailed representation, particularly when information is limited. More recently, the same idea has been discussed in the context of "reverse hierarchies" in the visual system (Fig 2.3), which puts the global level as the initial stage of representation, from which decomposition occurs (Hochstein & Ahissar, 2002). This scheme argues that an "early" visual area should be considered one which supports the earliest explicit representations, rather than one that is anatomically situated at the first cortical synapses.

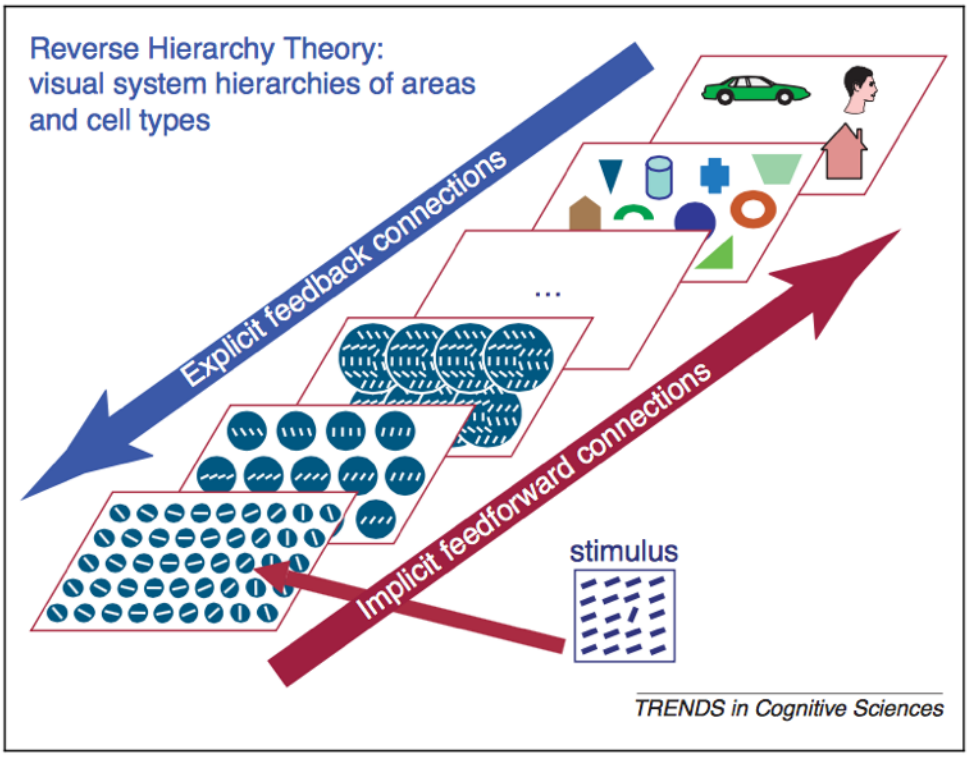


Fig 2.3. Reverse hierarchies in the visual system. *From Hochstein & Ahissar (2002). This model of visual representation proposes that information flowing in a feedforward manner (i.e. local-to-global) is accessible only implicitly, whilst the feedback connections (global-to-local) support the representations that guide our perceptual experience.*

The concept of a global level has been applied in work on scene categorisation to help explain how observers are able to rapidly extract the meaning from a complex visual scene without the time required to fully inspect the local features. Sometimes referred to as "gist" in this literature, observers can use the meaning of a scene to guide more search for specific objects, even when the images are viewed only briefly (Biederman, Rabinowitz, Glass, & Stacy, 1974). There is therefore a great deal of conceptual overlap between the Gestalt approach, the global/local distinction and the idea of representation through summary statistics of an input. However, the summary statistics approach provides a much more highly specified account of the transformation of information from features to whole. Representations via summary statistics can quantify the uncertainty associated with a

stimulus and can define the global level of perception as the best estimate of the central tendency of the set.

2.2. Explicit report of the summary measure in the absence of such knowledge about the component features.

Ariely (Ariely, 2001) addressed this issue by asking whether the ensemble representation is in some way distinct from the representation of the component parts. If he found such a dissociation, this would be evidence that particular prominence is given to the average, rather than it just being a property of the combination of already-held information. The task involved the presentation of a set of circles of different sizes (varying set size), followed by a single test circle. Two tasks could be performed with this test circle. The first task was "member identification", in which participants had to respond whether they had seen the single circle previously as a member of the set. The second task was to report whether the test spot was larger or smaller than the mean of the set ("mean discrimination"). The results clearly showed a difference in performance between the two tasks. Whilst participants performed well in the mean discrimination task, performance on member identification was at or close to chance in every condition. Ariely concluded that this was evidence for representation of ensemble properties (specifically the average here, although he also suggested that the similarity of the elements in a set influences the accuracy of the summary representation) in the absence of knowledge of the individual contributing members.

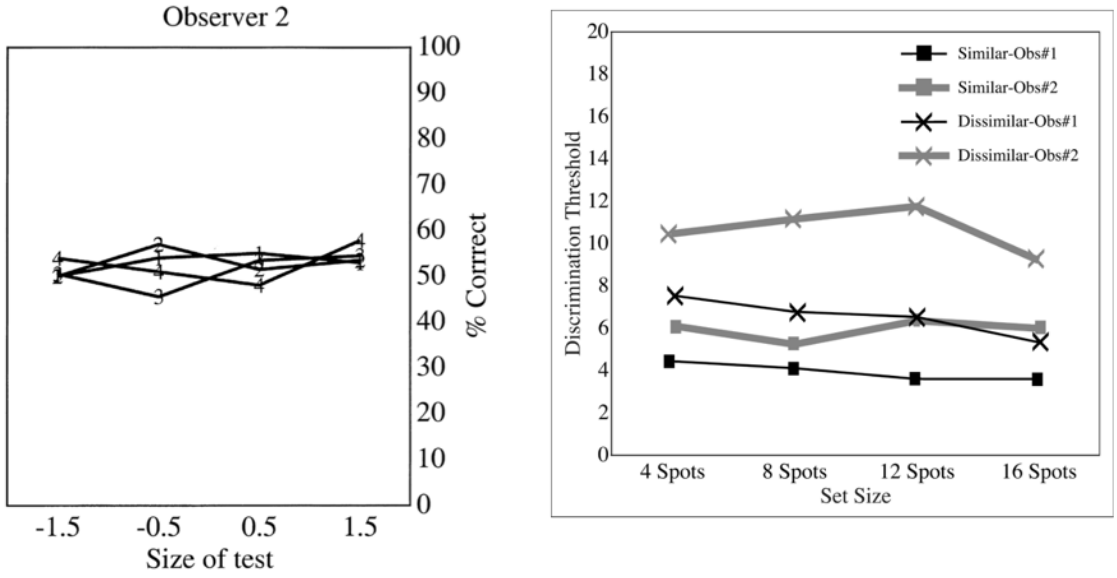


Fig 2.4. Results from Ariely (2001). *Left: data from a single observer for different set types (lines) on member identification test. Flat lines at 50% indicate the observer could not reliably report the identity of individual members of the set. Right: Discrimination thresholds for two observers for increasing set sizes. Flat lines indicate steady sensitivity based on the summary average across increasingly large set sizes.*

Chong and Treisman (Chong & Treisman, 2003) performed a series of studies using a comparison paradigm, in which participants were asked to report which of two sets had greater the greater mean size. They found that observers were just as accurate at making comparison judgements on the mean size of a set of heterogenous circles as they were at comparing the size of two single circles. This implies that averaging operates at the same level of accuracy as perception of a set of size one. Subsequent manipulations of exposure duration found that whilst general performance (on both the mean comparison and single item task) did improve with increasing exposure duration, at the shortest duration of 50ms, there was no difference in performance between the mean comparison and single items tasks. This shows that the averaging process can be completed very rapidly with minimal loss of information. Chong and Treisman argue that it suggests averaging must be an automatic, pre-attentive mechanism. In their view, averaging can be related to a form of parallel processing seen in visual search tasks (A. Treisman & Gormican, 1988; A. M. Treisman & Gelade, 1980), as opposed to a more serial form of processing in which each

item needs to be considered in series (this is proposed to be the level at which attention operates).

However, it need not necessarily be the case that the calculation of the ensemble is a process entirely divorced from that of member representation. Alvarez (Alvarez, 2011) highlighted that these results show only that the individual member representations are inaccessible for report, not that they are mutually exclusive. Whilst it would perhaps be a slightly extreme viewpoint that the featural and the ensemble representations are derived from entirely separate processes, it is necessary to clarify the extent to which they are separable. In other words, whether the ensemble representation is inextricably tied to the component features or whether it is a distinct quantity that can be used independently for alternative processing stages.

One of the few ensemble perception studies in the auditory systems demonstrated the statistical dependency of the summary representation. McDermott et al (McDermott, Schemitsch, & Simoncelli, 2013) tested the intuition that since a summary statistic represents a less noisy estimate of the typical value or pattern over time, a summary statistic representation should be able to discriminate patterns with different statistics better after prolonged exposure durations. They presented human listeners with auditory textures, which were formed from the combination of several simultaneous sounds. Since these textures were produced artificially, they could control the similarity of not only the local features but also the longer term statistical descriptions. Three example textures were presented, two of which were generated according to a single underlying summary statistic (but which differed in local features so they did not sound identical) and a third generated according to a different set of statistics. Participants were asked to identify which sound was generated from a different source. If listeners were sensitive to the statistical level description, then performance on this task should improve over time as the statistical descriptions of the similar sounds converged to their true value. This is exactly the pattern that was observed, suggesting that the auditory system, like the visual system, also builds a representation of summary statistics over time. As a further test of this

hypothesis, the converse was also tested, that is, whether it is more difficult to distinguish inputs with identical statistical descriptors based on their local features as sound duration increases. Again, the expected pattern was observed; listeners struggled to distinguish two physically identical stimuli from one with distinct local features but identical statistical identity. These results suggest that a summary representation based on averaging of inputs over time is formed without any direct instruction to do so and can be an accurate guide for discrimination.

2.3. Representation of summary statistics can be partially dissociated from representations reliant on selective attention.

Summary representations can take advantage of multiple sources of evidence to eliminate noise in an input and produce an accurate estimate of stimulus identity. This is similar to pooling models of sensory cortex, in which a large number of feature-selective neurons can be combined to form more complex representations (Freeman & Simoncelli, 2011). Summary representation can also be seen as a counterpart to a process such as selective attention, which forms a high-resolution but low capacity representation. One prediction from this view would be that changes in the deployment of selective attention would be far more damaging to accurate representation of local features than for accurate summary representations. Alvarez & Oliva (Alvarez & Oliva, 2008) asked participants to monitor the trajectory of multiple moving objects and report the number of times any object was in a particular zone. Simultaneously, a set of distractor elements were also moving randomly around the screen. This tracking task ensured participants were attending to a subset of presented stimuli. At the end of a trial, participants reported either the location of a missing single object or the location of the average position of a set of objects. When asked to perform the tracking task, localisation of individual items was better when the missing item had been part of the tracked (attended) set than when the missing item was from the distractor set. This suggests, as would be expected, that selective attention facilitated the encoding of individual item location. However, performance was equally as good for reports of average location for both attended vs distractor sets. This suggests that summary representations are not subject to the constraints of selective attention.

This result might suggest that summary representations are constructed entirely independently of the current level of attentional engagement. However, there is evidence to suggest that directing the visual system to focus on individual items may impair report of the average. Chong and Treisman (Chong & Treisman, 2005) linked the representation of individual members and the ensemble with different modes of attention, namely, focused and distributed attention. They used a dual task paradigm in which the primary task asked participants either to choose which of two circles was closest in size to the mean of a previously seen set, or to identify which of two circles had been a member of the previously seen set. The secondary task was introduced at the same time as the averaging set. Participants were asked to perform a visual search task for a singleton. The visual search was to find an O amongst Cs or a C amongst Os. Previous work had indicated that these were asymmetric searches - whilst a C among Os leads to the phenomenon of "pop-out", finding an O amongst Cs requires a much more directed, serial search approach (A. M. Treisman & Gelade, 1980). Chong and Treisman reasoned that these two searches require different attentional modes - a parallel search relies on distributed attention and a serial search requires focused attention. Thus, the member identification task should be performed better following serial search and averaging should be more accurate after a distributed mode of attention is cued via parallel search. This is exactly what was found. Summary representations and member identification may therefore represent two ends of a spectrum, with the preferred representation dictated by the task demands. However, the speed and accuracy of the extraction of summary representations might suggest that they are automatically calculated in response to complex environments.

Demeyere et al (Demeyere, Rzeskiewicz, Humphreys, & Humphreys, 2008) provide neuropsychological evidence that ensemble perception may rely on different mechanisms to the perception and identification of individual features. They describe patient GK, who developed simultanagnosia following bilateral temporoparietal strokes, with additional damage to occipitotemporal areas. Given that the critical feature of simultanagnosia is a very limited "attentional window" (often so limited that only one object can be processed

at any one time), the authors were surprised to find that GK could access some statistical information about a set of objects. GK was presented with a multielement display of coloured shapes followed by a probe item. He was asked to decide whether this probe item was new or had been part of the set. It was found that GK made more errors in rejecting the probe as new when the probe item was most similar to the average of the set of items. This interference shows that some representation of the average was extracted from the set, suggesting that summary representations do not rely on a functioning system for selective attention.

This interference effect also suggests that summary representations can bias perceptual judgements and perhaps then perception of an individual element is biased by a simultaneously computed average. One potential example of this kind of interactive relationship might be the Ebbinghaus illusion (Ebbinghaus & Dürr Ernst., 1902), in which the perceived size of a circle changes depending on the size of the surrounding circles (Fig 2.5). When the surrounding circles are larger, the reference circle also looks larger than an identical circle surrounded by smaller circles. One explanation of this illusion (and related examples (Howe & Purves, 2004, 2005; Howe, Yang, & Purves, 2005)) is that is occurs due to the probabilistic nature of the integration of perceptual features. In a display with larger circles, the average size is larger and therefore the size judgement of the reference is biased towards this average. This could therefore be an example of an interaction between perception of a single item or feature (size of a single circle) and the surrounding context, implying that the averaging process plays an active role in perception, beyond being an efficient representation of the individual elements.

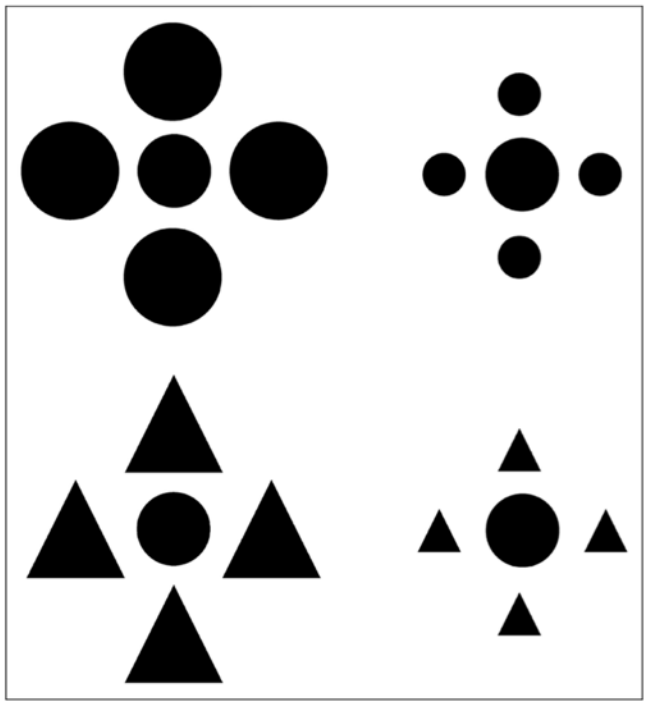


***Fig 2.5 Examples of the Ebbinghuas Illusion**. From Ariely (2001). The central circles are all equal size, but this percept is altered by the flanking circles. When the flanking circles are large, the central circle appears small. When the flankers are small, the circle appears large.*

2.4. Perceptual summary representation reflects real averaging strategies but does not necessarily rely on equally weighted items.

One question that arises from work of summary representation is whether it is possible for a statistical representation to reflect a perfect integration of information over time. Given the ubiquitous presence of capacity limitations in human cognition, it is unusual to find examples of such rapid and comprehensive information processing. If this is truly the case, it would perhaps strengthen claims that perceptual averaging reflects a qualitatively different form of visual representation. Much of the work described above suggests that perceptual averaging is a fast computation in which global, parallel processing ensures the contribution of all available samples to the summary representation. Given the well-described capacity limitations of human performance in all cognitive domains, it might, however, be reasonable to expect some limitations on the ability to calculate summary representations. Still, the ability of human observers to accurately respond on the basis of summary measures suggests that the brain has found a mechanism to overcome capacity limitations without any significant information loss.

In the search for an alternative mechanism, Myczek & Simons (Myczek & Simons, 2008) suggested that rather than producing a true estimate based on all available samples, the brain implements a sampling strategy such that only a subset of samples contribute to the final summary judgement. This approach suggests that we use a focused-attention strategy that can appear to mimic a parallel strategy, although they argue that this may only apply for averaging of features for which we do not have early visual representation, such as for averaging of absolute size as opposed to motion or orientation. In support of this view, they simulated the performance of an ideal observer who implements various sampling strategies. Using the same size averaging paradigm as Ariely (2001), they compared the performance of real observers to that predicted by ideal observers who based their decision on either 1, 2 or 3 randomly selected components. They found that the performance of real participants could be well approximated by a subsampling strategy based on one or two component samples. This appears, then, to indicate that instead of relying on a representation of the distribution of inputs, participants select a few elements at random according to a focused attention strategy.

However, many objections have been raised in response to this finding. Firstly, the model assumes that there is no noise at the stage of feature encoding, a perhaps unrealistic assumption. Ariely (Ariely, 2008) questioned whether this model would be able to recreate the second feature of his original averaging task - the lack of recognition performance for individual members. If participants were using a subsampling strategy through focused attention, then they should be able to subsequently identify at least those members of the set. For small set sizes, this would predict performance above chance for member identification, which was shown not to be the case by Ariely's 2001 experiment (Ariely, 2001). Furthermore, the explanation advocated by Myczek and Simons cannot account for examples of perceptual averaging that do not have an obvious neural correlate for the low level features, such as texture perception (Gorea, 1995).

Chong et al (Chong, Joo, Emmanouil, & Treisman, 2008) presented new findings (in addition to several conceptual criticisms) in defence of a parallel mechanisms for summary

perception. One experiment directly implemented the results simulated by Myczek & Simons by testing the accuracy of an average derived from a full set (numerous items) or from a subsample of elements. Participants were asked to judge which of two sets had a larger mean. The sets were either the full set or represented a subsample from a full (but unseen) set. The simulation data would suggest that these two conditions would be equivalent if subjects were basing average judgements on a subsample. Accuracy was, however, significantly higher when participants made their judgements based on the full set, rather than on a subsample. Furthermore, the accuracy of judgements based on the samples was less than would be expected that if participants were perfectly integrating the evidence.

These results therefore suggest that human performance is limited by more than just attentional demands. The simulations of Myczek & Simons did not take this into account, and found that human performance could easily be matched by an ideal observer subsampling with perfect integration of those elements. It seems therefore that a subsampling strategy cannot account for all the behavioural findings and that human observers can form accurate summary statistic representations. However, even if an average judgement does rely on a parallel integration across all available samples, it need not necessarily imply that all elements contribute equally to the overall decision. This proposal has found some support. De Gardelle & Summerfield (de Gardelle & Summerfield, 2011) asked participants to form average feature judgements on the basis on eight individual elements, sampled from a Gaussian distribution. Logistic regression was used to quantify the extent to which each element contributed to the response on a trial-by-trial basis. This analysis method showed that whilst each element did contribute, elements that fell at the centre of the trial distribution contributed the most, whilst elements towards the tails of the distribution were given less weight (Fig 2.6). Such a mechanism can help make a decision process robust to noise fluctuations, acting like an analyst removing outliers from data. Additionally, the downweighting according to position in the trial-to-trial sample distributions suggests that the observers are encoding the distribution of information, rather than just a single point estimate of the mean.

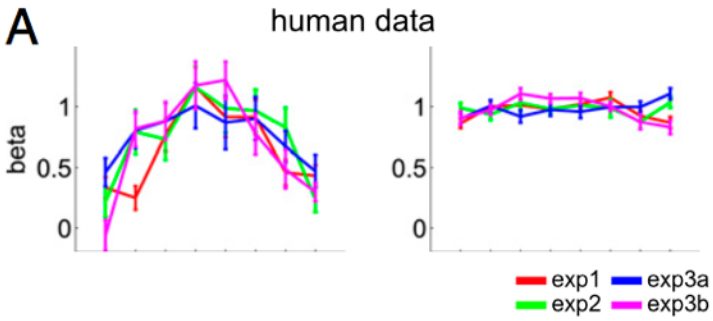


Fig 2.6. Weighting functions from de Gardelle & Summerfield (2011). Regression weights (betas) are plotted for the ranks of elements along a task-relevant (left) or task-irrelevant (right) stimulus dimension. Large betas indicate the position of those elements that were weighted most strongly in the observers' choices.

Similar evidence of downweighting might also, however, reflect capacity limitations. In a task which required observers to average the orientation of sequentially presented Gabor patches, Wyart et al (Wyart, de Gardelle, Scholl, & Summerfield, 2012) showed that the extent to which any given patch contributed to the average (i.e. its weight) varied over time. Specifically, the impact of patch was related to the phase of ongoing delta oscillations. When the arrival of a patch coincided with the peak of a cycle, the patch was observed to have a greater impact on the eventual average judgement that if that patch had coincided with the trough. Here, the downweighted samples did not reflect information falling in a particular portion of feature space but instead reflected information falling at a particular temporal location. This may still help to create a robust representation by allowing the processing of one element to proceed without interference from the next stimulus, if the presentation rate exceeds the processing capacity of the visual system. Therefore, although observers are able to report accurate estimates of the average, the visual system may employ mechanisms that ensure this process is as noise-free as possible, even if this means disregarding information over time.

2.5. Summary statistics act as cues for identity in noisy or low-resolution environments.

The robust, or stable, nature of these summary representations might make them particularly useful in noisy environments. Given the constant small fluctuations in visual input from the environment, the brain may choose a form of representation that is the most stable over time. Summary representation would be one way to achieve this stability. The phenomenon of change blindness may be a case in which this kind of stable representation is chosen over one that is sensitive to all the local fluctuations in input. It has been repeatedly shown that human observers are very bad at identifying local changes in complex scenes (Simons & Levin, 1997) , even if they are attending at the location of the change (Ballard, Hayhoe, & Pelz, 1995; Simons, 1996). One explanation for change blindness is simply that we cannot process all objects within a scene simultaneously and therefore have to engage in something like a serial search process to find the changing location. However, this explanation then struggles to account for findings in which all elements change at the same time but the change blindness phenomenon persists (Suchow & Alvarez, 2011). Ensemble representations could contribute to an alternative explanation. If an observer is monitoring only the ensemble representation, then changes to local elements will only be detected if they introduce a significant perturbation in the ensemble. Saiki & Holcome (Saiki & Holcombe, 2012) carried out a change blindness experiment in which they considered both the global statistical information and the local featural changes. Using a random dot motion task, they showed that changes in local features went unnoticed only when the change did not result in an overall change in stimulus statistics.

A similar conclusion was drawn by Haberman and Whitney (Haberman & Whitney, 2011) in a task using sets of emotional faces. They asked participants to look at two sets of faces, each with a different emotion. In the change condition, four faces switched emotion (e.g

from angry to happy). When participants were asked to detect the locations which showed a change in emotion, their performance was very poor, but they could accurately perform a comparison on the average emotion of the two sets. Therefore, the individual items did contribute to an ensemble representation but were not individually accessible for report, or at a level strong enough to guide choice behaviour.

When perceptual representation is too limited to identify individual features, the level of summary representation may again be useful for providing an estimate of identity; many faint and blurry images can average to produce a single, clearer image. It has been proposed that this may be a contributing mechanism to a second visual phenomenon - crowding. In crowding experiments, identification of a peripheral target is severely impaired by adding distractors around the target, compared to a condition in which the peripheral target appears alone (for a review, see (Levi, 2008)). Parkes et al (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001) showed that observers can produce an accurate estimate of the orientation of a set of objects, even when a single item in the set is not reported accurately due to crowding. This could be due to an automatic averaging that produces a summary representation at the expense of item-level representation. In neural terms, this might be expected to be due to automatic pooling of low level neurons with small RFs onto downstream neurons with RFs large enough to encompass the entire visual array. Note that this pooling model would assume that the neurons with the most influence on behavioural response would be those with the RF size most suited to the current stimulus input. It should also be noted that an explanation of crowding based on automatic averaging cannot completely account for the phenomenon, since the spatial arrangement of flankers around a target can have a substantial influence, even when the average remains constant (Livne & Sagi, 2007).

The role of ensemble representation in crowding received a more explicit test from Balas et al (Balas, Nakano, & Rosenholtz, 2009). Here, the authors created artificial stimuli (“mongrels”) that had the same statistical structure (as defined by a texture synthesis algorithm, which conserves some of the low level features) as a set of recognisable letter

stimuli but had the appearance of a jumbled set of features. Participants performed a sorting task with the mongrel stimuli in which they were asked to sort them into sets according to which one of a possible four target stimuli they thought was the source for the "jumbled" image. They subsequently completed a battery of crowding tasks. It was found that performance on the mongrel sorting accurately predicted performance on the crowding tasks, suggesting that the ability to distinguish or match images on the basis of their statistical similarities might be related to performance limitations on the crowding task. Whilst being far from conclusive evidence, this approach suggests that the interactions between stimuli may be an important contributing factor for all kinds of visual tasks.

Whilst it may be the case that summary statistics allow for robust and stable representations, it might also be the case that these summary representations could be used to identify outlying elements that do not belong to the current set. This is exactly the computation required to successfully complete visual search tasks. However, it would depend on the summary representation encoding not just an estimate of the average, but encoding the full distribution of inputs across the relevant feature space.

2.6. To what extent is a full summary stimulus distribution represented in the brain?

Although it seems clear that human observers can easily extract and use the ensemble information in a display, it is unclear to what extent this ensemble information reflects the underlying statistical properties of the display. Evidence for sensitivity to the statistics of the display would require demonstration the visual system is sensitive to the distribution of information.

Abundant evidence has already been presented above to show that the human visual system can calculate the mean of a set of inputs. Chong & Treisman (Chong & Treisman, 2003) showed that the form of the distribution of element value in a set could modulate

the judgement of the average. However, if neural systems are able to represent information as probability distributions then there is potentially more information available than just the mean. As already discussed, the variance associated with a stimulus or feature information can determine to what extent a cue drives behaviour. The full form of the input distribution may therefore be of interest to perceptual systems. In Ariely's 2001 original experiments on size averaging, a manipulation of set heterogeneity (variance) was also included. He found that participants were better at classifying the test circle with respect to the set average when that set was more homogeneous.

This latter example strongly suggests that the quality of a perceptual judgement can be influenced by several statistical features of the input. The relationship between these multiple influences on perceptual judgement, however, has not yet been fully explored. In the field of value-guided decision making uncertainty, it has long been accepted that there are multiple, independent, contributing factors which together determine the difficulty of a decision. Decisions become more difficult when uncertainty increases, but uncertainty can come in several forms. For example, the value of an option can be seen as one source of uncertainty (Jocham, Hunt, Near, & Behrens, 2012; Philiastides, Biele, & Heekeren, 2010; Rushworth, Noonan, Boorman, Walton, & Behrens, 2011). Additionally, the risk (defined as the variance associated with the outcome) associated with an option adds another, independent, source of uncertainty (O'Neill & Schultz, 2010). It remains to be seen, however, whether a similar scheme of dissociable sources of decision uncertainty can be applied to perceptual systems. Although the outcomes of perceptual and value-guided decisions can be very different, it could be that similar guiding principles underlie the way the decisions are made (Summerfield & Tsetsos, 2012).

Thus, although several studies indicate that the overall shape of the input distribution can influence the speed and accuracy of perceptual decisions it is unclear to what extent the parameters of these distributions can be classed as independent sources of decision uncertainty. For example, although decades of experiments using the random dot motion task have shown that increasing the strength of the evidence through changing the

coherence of motion improves accuracy and response times, it is not clear whether this is due to a change in the mean (average number of dots moving in the same direction) or to the change in variance. In this task, changes in the mean and variance are confounded as increasing the coherence will decrease the variance. De Gardelle & Summerfield (de Gardelle & Summerfield, 2011) created a task in which the mean and variance of a set of items were entirely decorrelated. The stimuli could be described based on two dimensions, shape (blue or red) or colour (square or circle). Participants were asked to make a category judgement on the basis of the average feature value (shape or colour) of a set of eight such stimuli. These sets of "squircles" were drawn from a Gaussian distribution defined by a particular level of mean distance from category boundary and standard deviation.

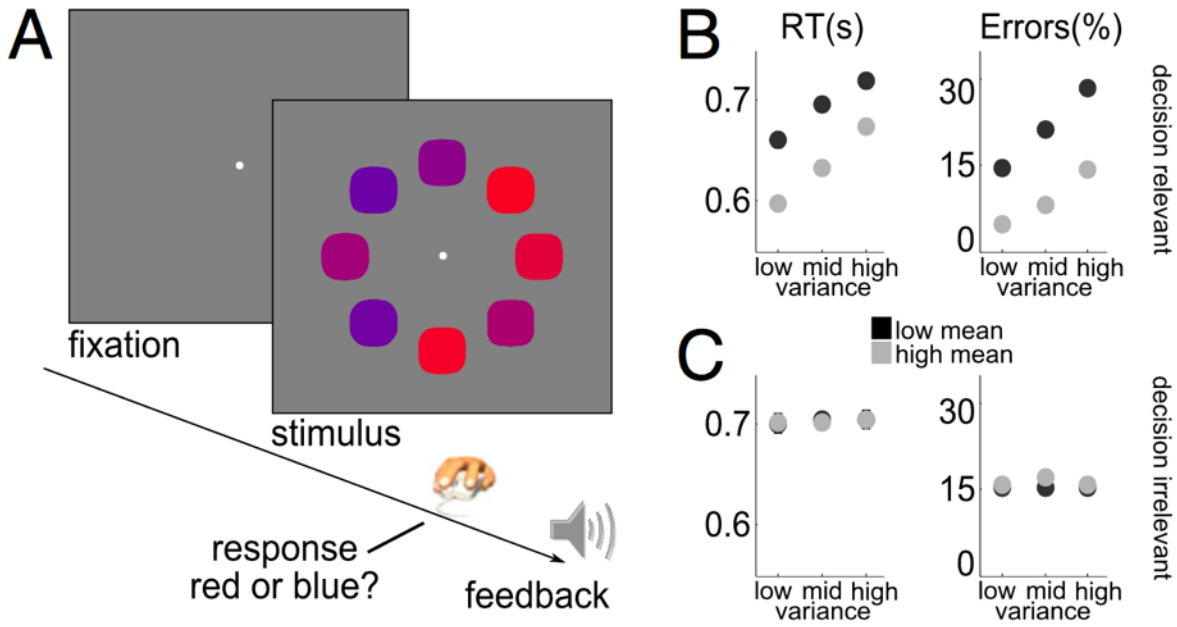


Fig 2.7. Task and behavioural results from de Gardelle & Summerfield (2011). Left: Observers were asked to perform a multielement averaging task, in which they should classify a set of coloured shapes on the basis of (here) their average colour. The mean distance to category boundary of the task-relevant information and the variability of this information were orthogonalised across trials. Right: Behavioural results (response times, RTs and errors) for relevant dimension (top row) and irrelevant dimension (bottom row).

Participants in this task showed increased reaction times (RTs) and decreased accuracy when the mean of the set was closer the category boundary, as would be expected from theoretical models of categorisation, such as signal detection theory (see Chapter 1). However, since the mean and variance were orthogonal by design in this experiment, the contribution of the set variance could also be measured. It was found that increasing variability was also associated with reduced performance (slower RTs and increased errors), which was an effect that showed no interaction with the level of mean distance from category boundary. Therefore, participants in this task were sensitive to two independent sources of uncertainty - that induced by distance from category boundary and that due to increasing spread of information. Sensitivity to the variance also shows that the level of summary representation extends beyond simply a single estimate of the mean.

2.7. Chapter Summary

In summary, there is a large amount of evidence supporting the idea of summary representations in the human visual system. These representations can persist even in the absence of any reportable knowledge about the constituent features. Representation at the summary level perhaps provides an alternative method for avoiding capacity limitations and can provide a robust representation in the presence of noisy input. The work presented in the following chapters aims to extent these ideas using a multi element averaging task which allows for orthogonalisation of the mean and variability of visual information.

Chapter 3: fMRI Evidence for Dissociable Sources of Perceptual Uncertainty

This chapter uses fMRI to investigate the neural overlap in brain regions sensitive to two dissociable sources of perceptual uncertainty: the overall mean distance of the evidence from the category boundary and the variability of this information. It also aims to investigate whether a model assuming probabilistic representation can account for neural activity, as measured by the BOLD signal.

As argued in the introductory chapters, there is mounting evidence to support the idea of the representation of visual information at the level of its summary statistics, as opposed to the representation of each individual feature (Alvarez, 2011; Ariely, 2001). This form of representation can provide the information necessary for categorisation when multiple sources of evidence need to be combined. However, how this summary representation arises is still debated. In the context of categorisation decision processes, these summary representations could reflect the final stage of a process in which evidence is accumulated. Work on perceptual decision processes has suggested a general framework in which the available evidence is distilled into a single value that changes over time to reflect the evidence in favour of one response option or the other – the decision variable. Although the exact form of accumulation is debated, there is broad support for the general mechanism from behavioural modelling (Ratcliff & McKoon, 2008), single unit recordings (Gold & Shadlen, 2007) and human neuroimaging (Heekeren et al., 2008).

Despite the broad consensus on the idea of evidence accumulation for perceptual categorisation the specifics of implementation are far from clear. One important issue is the identification of the relevant summary quantity that underlies the decision variable. Although the mean of the signal will generally be the most informative for categorisation tasks, many summary measures can be computed from a set of inputs, although little attention has been given to systematically controlling higher order statistics of the perceptual evidence. A commonly used task asks observers to identify the average direction of motion of a group of moving dots (Shadlen & Newsome, 2001; Williams &

Sekuler, 1984). The coherence of the motion can be manipulated, so that under 100% coherence all dots move in the same direction, whilst under 0% coherence all dots move independently in a random direction. Sequential sampling models all predict that the time taken to reach a decision threshold will be increased as the level of stimulus noise increases. The coherence level of the stimulus, or signal to noise ratio, can therefore be seen as a single measure that will determine decision latencies and accuracy.

However, this single measure confounds two potentially independent summary measures of evidence quality: the mean signal strength and the variability of the information. As the motion of the dots becomes more coherent, the strength of the evidence will increase but the total variability of motion will decrease. Therefore, it is difficult to determine whether performance is facilitated purely by an increase in the average number of coherent dots (evidence strength) or by less variability in the samples evidence (or some combination of the two).

To address this issue, de Gardelle & Summerfield (de Gardelle & Summerfield, 2011) designed a stimulus set in which the evidence strength was orthogonal to the reliability. Observers were asked to categorise a set of coloured shapes on the basis of either their average colour or average shape. Evidence strength was manipulated by changing the distance between the average feature value and the category boundary. As discussed in the introduction, moving the average close to the boundary negatively impacts categorisation performance. The orthogonal manipulation of evidence reliability was implemented by changing the variability (specifically, the standard deviation of feature values in the stimulus array) of the presented evidence. Thus, any given array could be “low” (close to the boundary) or “high” (far from the boundary) mean with “high” or “low” variability.

Although the effects of changing mean distance to category boundary have been long-reported, it was unclear exactly what should be predicted about the influence of feature variability. Intuitively, increasing the variability (decreasing the reliability of the signal) should make any decision more difficult as the observer may wish to delay response and collect more evidence to over come the increase in noise. However, several popular models of perceptual decision processes predicted that the fastest response times should be

associated with the highest levels of variability. This happens (in the case of the diffusion model) because the greater variability drives the decision variable to threshold through noise earlier in the decision process than when the variability is lower. The behavioural performance of participants clearly showed, however, that the trial mean distance to category boundary and the variability were independent contributors to response accuracy and latency. Both increased variability and decreased distance to category boundary resulted in slower and less accurate responses. Therefore, the decision variable must be a composite measure which takes into account several sources of uncertainty.

De Gardelle & Summerfield presented an alternative model (Log Probability Ratio, or LPR model) that was able to account for several features of their behavioural data. This model assumes that the decision process involves accumulation of the probabilistic values associated with each point in feature space, rather than directly accumulating the individual feature values of the contributing evidence. Representation of information in such a probabilistic space has been proposed to account for performance on tasks requiring the integration of multiple sources of evidence (Ernst & Banks, 2002; Kira et al., 2015; Yang & Shadlen, 2007), and several schemes of neural coding have provided plausible mechanisms for such transformations (Ma & Jazayeri, 2014).

In learning this transformation from sensory to probabilistic (“decision”) space, the transfer function takes on a sigmoidal shape. This form in not assumed a priori or imposed artificially, but instead arises from the sequence of arrays with intermixed levels of mean distance to boundary and variability. Nonlinear representation of sensory spaces appears to be a common feature of neural coding schemes (Reynolds & Heeger, 2009; Schwartz & Simoncelli, 2001). One consequence of the LPR transformation is that values at the extreme ends of the sensory space (referred to as outliers) saturate in the extent to which they contribute information to the decision. That is, after a certain point in the sensory space, all elements predict the response with the same probability. This feature of the LPR model predicts that there should be a corresponding down weighting of these elements in the decision process - a prediction that was confirmed empirically.

The LPR model also provides a parsimonious account for the independent effects of mean and variance on behaviour. In this model, the single transformation to decision space accounts for both effects; low mean trials will have a lower mean LPR across all elements, equivalent to reducing the strength of the input. Trials with high variability will be disproportionally penalised by the downweighting of outliers, so that very predictive elements contribute less evidence than they should to the decision. In an accumulation to bound decision process, this will decrease the drift rate (rate of evidence accumulation) and result in prolonged response times.

This single process feature of the LPR model opens the question of whether there is also a single neural correlate of such a transformation, or whether there is a broader representation of decision-relevant evidence at the summary statistic level. Recent human fMRI studies have highlighted that a wide range of brain regions respond to uncertainty in decision relevant information (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Grinband, Hirsch, & Ferrera, 2006; Huettel, Song, & McCarthy, 2005). Sensory cortices seem to be most related to the representation of sensory evidence, with the specific region activated dependent on the stimulus class to be categorised (Ho et al., 2009). Thus, dot motion task performance activates medial-temporal areas (Britten et al., 1996; Ho et al., 2009; Huk & Shadlen, 2005), whilst face-house decisions activate fusiform or parahippocampal regions respectively (Heekeren et al., 2004). Interference with these regions produces a bias identical to what would be expected by changing the evidence (Huk & Shadlen, 2005). Other (perhaps parallel, rather than sequential (Heekeren et al., 2008)) stages of the decision process, such as performance monitoring or accumulation of the decision variable have been associated with broader networks extending into prefrontal and parietal association cortices (Gold & Shadlen, 2007; Huettel et al., 2005). It is unclear, however, whether these networks are equivalently modulated by dissociable sources of perceptual uncertainty.

This project therefore primarily aimed to investigate, using fMRI, whether orthogonal manipulation of the mean and variability of decision relevant information would be reflected in distinct neural correlates of decision uncertainty. Furthermore, we seek to

further validate the LPR model by testing whether brain regions represent information in the transformed, probabilistic, decision space.

3.1 Methods

3.1.1. Participants. Twenty right-handed volunteers (reporting normal or corrected-normal vision and no history of neurological problems), aged between 20 and 35 (9 females, 11 males), provided informed consent and were paid £30 compensation for taking part. The study was approved by local ethics committees.

3.1.2. Stimuli.

Stimuli were created and displayed using PsychToolBox (www.psychtoolbox.org) for MATLAB (Mathworks). Stimuli were presented on a custom shielded Samsung 40" LCD screen (LTA400HF1) at a distance of 240cm. On each trial participants viewed an array of eight elements (coloured shapes) circularly arranged (radius ~3° visual arc) around a white central fixation point (5 pixels radius). Elements were equally spaced, equiluminant and covered an equal area on the screen (width = height = 50 pixels for pure circle). Stimuli were presented on an equiluminant grey background. Each element was defined by a shape parameter (S) that determined its position on a continuous transition between a square (S=-1) and a circle (S=+1), and a colour parameter (C) that determined its position on a continuous transition between blue (C=-1) and red (C=+1). This is described in detail elsewhere (de Gardelle & Summerfield, 2011). On each trial, the parameter values for each dimension were drawn independently from a Gaussian distribution with mean μ and standard deviation σ. To ensure equal precision of the mean in all conditions, resampling occurred until the sampled trial μ and σ fell within a tolerance of 0.1% of the desired values.

3.1.3. Design.

The mean μ could take one of four values: two either side of the category boundary, giving rise to two absolute distances to category boundary, $|\mu|$, which we refer to as low-mean

vs. high-mean conditions. The standard deviation σ was manipulated in three levels. The shape and colour dimension were manipulated independently within and across trials. The task involved only one dimension (either colour or shape), varied across different blocks, such that each dimension could in turn be relevant or irrelevant for the decision. This afforded us a 2 (relevant |μ|) x 3 (relevant σ) x 2 (irrelevant |μ|) x 3 (irrelevant σ) within-participant factorial design.

3.1.4. Thresholding.

To equalize difficulty across participants and shape and colour tasks we used an adaptive procedure, in which the mean parameter of the array was varied to achieve an accuracy of 75% (low-mean condition) or 85% (high-mean condition). The three levels of variance were identical for all participants (0.1, 0.2 and 0.3). Each participant completed this staircase procedure (4 blocks: low and high mean condition in both tasks, with 144 trials in each block) on a standard testing PC on a day prior to the scanning session.

3.1.5. Task and Procedure.

On each trial, observers classified a circular array of eight elements ('squircles') according to their average colour or shape (Fig 3.1a). As described above, each of the eight elements took on a colour value (red to blue) and shape value (square to circle), both parameterised in the range -1 to 1, with the category boundary falling at zero (Fig 3.1b). Colour and shape were deemed decision-relevant in alternating blocks and the observers' task was thus to respond circle/square or red/blue according to whether the average feature value on the *relevant* dimension was greater or less than zero (ignoring the *irrelevant* dimension).

As shown in Fig 3.1a, the stimulus array appeared 500 ms after the onset of central fixation point, and remained on the screen for 1500 ms, during which time participants judged the average shape or colour of the elements in the array, depending on the task

block. Responses were made by pressing one of two keys on a button box (scanner) or computer mouse (thresholding task). In the scanner, participants used the index fingers of both hands to make responses whilst for thresholding the index and middle fingers of the same hand were used. Response mappings were fully counterbalanced across both dimensions between participants. At array offset, auditory feedback indicated response accuracy on each trial. Two ascending tones (400-800Hz, 100ms each) indicated a correct response whilst two descending tones (800-400Hz, 100ms each) were given for incorrect responses or misses (no-response trials). In the scanner, a jitter of 4+/- 2s was introduced into the interstimulus interval (ISI); for the thresholding task, the ISI was jittered uniformly around 1s (min 0.85s, max 1.15s).

Each block began with an instruction screen indicating the relevant decision dimension (either shape or colour) and the response mapping for the block. Response mappings were fully counterbalanced across participants. Participants underwent two blocks of 144 trials for each task, for a total of 576 trials, and the order of tasks was pseudorandomised.

3.1.6. Eyetracker: acquisition. An MRI compatible eyetracker (Eyelink 1000 tracking device - SR Research, Ontario, CA) was used to monitor eye movements in the fMRI scanner. The tracker was adjusted before each experimental block with a calibration/ validation procedure in which participants followed with their eyes a small circle moving between 9 locations on the screen. Due to the difficulties of eyetracking in a scanner environment, accurate calibrations were not always possible and so eye data is presented from 10 participants. Data were collected using the PsychToolBox Eyelink toolbox and analysed using in-house customised Matlab scripts. After downsampling to 200 Hz, we calculated the mean and variance of the displacement of the eye on each sample. ANOVAs were then used to compare these estimates for different levels of stimulus array mean and variance.

3.1.7. Behavioural analyses. For each participant, we calculated accuracy (percent correct) and response latencies on the correct choices, in the 3 relevant variance x 2 relevant mean conditions of our design, and carried analyses of variance (ANOVA) at the group level (Fig 3.1d). The same procedure was used to analyse the effects of the irrelevant dimension.

We also calculated a weighting profile across elements (Fig 3.1c), as in our previous work (de Gardelle & Summerfield, 2011). This weighting profile is a plot of regression weights describing the contribution of each element to the trial decision. For each participant, we fit a probit regression in which the weighted sum of the 8 relevant feature values plus a constant term predicted the probability of "positive" choices, on a trial-by-trial basis. We sorted the 8 values in each trial before including them as predictors, so that the weights (i.e. the regression coefficients) corresponded to the different ranks along the task-relevant dimension. Then, we divided all weights in each participant by their root mean square, a normalization procedure that minimized the influence of unreliable estimates and that was neutral with respect to the weighting profile. We then compared the average normalized weights for the 4 outlying elements (i.e. the elements ranked 1, 2, 7, 8 in the trial) and the 4 inlying elements (ranked 3, 4, 5, 6), in a paired t-test across participants.

3.1.8. Computational model. As in our previous experiments (de Gardelle & Summerfield, 2011), we calculated for each participant and task the proportion of trials in which a particular colour or shape value x was associated via feedback with the left-hand response category ($P(L|x)$) or the right-hand response category ($P(R|x)$). Taking the logarithm of the ratio between these probabilities, we transformed the stimulus value x into a LPR value (log probability ratio; eq 1). This LPR value quantifies the association between x and the two response categories

$$LPR(x) = \log\left(\frac{p(R|x)}{p(L|x)}\right) \qquad \text{(eq 1)}$$

Here, we calculated these probabilities in 10 bins along the x axis (each bin contained 10% of the data), and fit them with a sigmoidal function. From then, we simulated a diffusion model in which the mean LPR over the 8 elements was used to drive the accumulation of evidence towards choice (see de Gardelle & Summerfield, 2011). The diffusion model had two free parameters for each participant: the noise in the diffusion and the amount of accumulated evidence needed to trigger the response. These parameters were optimized for the model to fit the 6 error rates for each participant and task. The simulated RTs (calculated in cycles) were then linearly scaled to the range of human RTs (in seconds), by setting the cycle duration such that the human and simulated RTs had the same standard deviation across all trials, and by adding a constant offset such that the simulated and human RTs had the same mean across all trials. This rescaling ensured that we could compare simulated and human RTs, while being neutral with respect to the profile of RTs across conditions.

3.1.9. fMRI acquisition and preprocessing. Images were acquired in a 3 Tesla Siemens TRIO with a 32 –channel head coil using a standard echo-planar imaging (EPI) sequence. Images were 64x64x36 volumes with voxel size 3x3x3 mm; acquired with a 2s repetition time (TR) and 30ms echo time. Four runs of 412 volumes were obtained, each of which lasted approximately 15 minutes and corresponded to one experimental block of 144 trials.

Preprocessing of the imaging data included correction for head motion and slice acquisition timing, followed by spatial normalization to the standard template brain of the Montreal Neurological Institute (MNI brain). Images were resampled to 3mm cubic voxels and

spatially smoothed with a 10mm full width at half-maximum isotropic Gaussian kernel. A 256s temporal high-pass filter was applied in order to exclude low-frequency artifacts. Temporal correlations were estimated using restricted maximum likelihood estimates of variance components using a first-order autoregressive model. The resulting nonsphericity was used to form maximum likelihood estimates of the activations.

3.1.10. fMRI analyses. All fMRI analyses were carried out using SPM8. SPM orthogonalises regressors by default, but we ensured that this feature was turned off. We analysed the data in two distinct ways. In 'native space' analyses, we created independent regressors encoding the parameters |μ| and σ of the stimulus array. In 'decision space' analyses we substitute these for their counterpart in terms of the LPR-transformed feature values for each array, which we denote U_{Mr} and U_{Vr} (the subscript r indicates that this is about the task-relevant dimension). The decision space values represent the output from a proposed stage of processing in which the feature values (in colour or shape space) are passed through a sigmoidal function. Including this transformation accounts for the behavioural finding of a weighting function in which outlying elements are downweighted compared to those at the centre of the trial distribution of feature values.

$$mLPR = \frac{1}{8} \times \sum_{k=1}^{8} LPR(x_k) \qquad \text{(eq 2)}$$

$$vLPR = \frac{1}{8} \times \sum_{k=1}^{8} \left(LPR(x_k) - mLPR\right)^2 \qquad \text{(eq 3)}$$

$$U_{Mr} = -\left|mLPR\right| \qquad \text{(eq 4)}$$

$$U_{Vr} = \sqrt{vLPR} \qquad \text{(eq 5)}$$

We calculated the decision space values as described in eq2-5. First, we calculate the mean (mLPR) and variance (vLPR) of the LPR over the 8 elements (eq 2 and 3). Then, we define U_{Mr} and U_{Vr} (eqs 4 & 5) such that both quantities positively scale with the intrinsic difficulty of the stimulus in the categorization task, which increases when the mean evidence approaches zero or when the feature values become more variable. These quantities were then used as predictors for the BOLD responses. In a separate analysis, we substituted the variability of the evidence for a different regressor encoding the sum across elements of the absolute value of the LPR.

Both native space and decision space analyses included the mean and variance regressors for the relevant and irrelevant dimensions (1-4), as well as separate regressors encoding (5) the feedback coded positive for correct and negative for incorrect trials, and nuisance regressors encoding movement parameters estimated from the realignment phase (7-12). These analyses were carried out independently for each participant, and the resulting T-statistics for each regressor were then subjected to T-tests at the group level. Voxels reported are those that survived at an uncorrected threshold of $p < 0.0001$. Full details of these voxels can be found in Tables 1 and 2.

In order to plot activity at different regions for the relevant and irrelevant mean and variance, we placed a sphere of 5mm radius on the peak voxel in each cluster identified as responding to either the relevant mean or relevant variance. We used a spherical ROI centred on the peak voxel, rather than functionally-defined ROI, to avoid having to select a statistical threshold above which to include voxels. We then plotted its response to relevant and irrelevant mean and variance. We report statistics for (only) the response within this region to the orthogonal factor, i.e. to the mean when data were extracted on the basis of the whole-brain search for voxels responsive to variance, and the variance when data were extracted on the basis of the whole-brain search for voxels responsive to mean. Because the regressors for the mean and variance (either $|\mu|$ and σ or the mean and standard deviation of LPR values) were orthogonal by design, this approach avoids any circularity or 'double-dipping'.

In order to show haemodynamic response functions (Figure 3.3d), a further whole brain finite impulse response (FIR) analysis was run with the native-space values. This model replicated the structure of the main analyses, including the mean and variance of both the relevant and irrelevant dimension. In this analysis, 16 time bins (each corresponding to 2s) were entered into the design matrix for each regressor described above. Positive and negative feedback, as well as movement parameters, were also included in the design matrix for this analysis. Data in Figure 3.3d are presented from a sphere of 5mm radius centred on the peak voxel within the dmPFC ROI.

3.2 Results

3.2.1. Behavioural data.

Pre-experimental calibration ensured comparable performance in shape and colour tasks for the fMRI experiment (see methods). Mean error rates (shape: 18%, colour: 16%) and response times (shape: 800ms, colour: 803ms) were not different across tasks. ANOVAs revealed main effects of both mean (i.e. proximity to category boundary) [$F_{(1,20)} = 92.7$, $p < 0.001$] and variance (i.e. element heterogeneity) [$F_{(2,40)} = 18.0$, $p < 0.001$] on correct response latencies (Figure 3.1d, top left panel), but no significant interaction [$F_{(2,40)} = 0.46$, $p < 0.58$]. Comparable effects were found for accuracy (Figure 3.1d, bottom left panel), with more errors occurring on trials with low mean than high mean [$F_{(1,20)} = 104$, $p < 0.001$], or trials with high variance vs. low variance [$F_{(2,40)} = 7.80$, $p < 0.001$], and no interaction [$F_{(2,40)} = 2.70$, $p < 0.15$]. The irrelevant mean and variance had no significant effect on RT or errors (all p-values > 0.05; Figure 3.1d, right panels). Analyses of eye movements (mean length and variance of the eye path) confirmed that they did not differ between conditions (all $p > 0.19$, Fig A1.1). Together, these data confirm the previous finding that array mean and variance have independent (i.e. non-interacting) effects on behaviour (de Gardelle & Summerfield, 2011).

When decisions involve integration of multiple independent sources of evidence, the question additionally arises of how much each source impacts the final choice. As described above, de Gardelle & Summerfield have previously reported that the most extreme evidence (i.e. evidence that falls closest to the tails of the trial distribution of feature values) on any given trial will be downweighted relative to the evidence falling at the centre of the trial distribution. Calculating the weight that each element (sorted by its rank) carried in the choice using logistic regression (see methods), we replicated this finding here. We observed significantly higher weights (regression coefficients) for inlying vs. outlying items [$t_{(20)}=2.72$, $p < 0.013$], when collapsing across relevant feature dimension (colour or shape). These data are shown independently for relevant feature dimensions in Figure 3.1c (left panel).

3.2.2. Computational model.

It has previously shown that it is possible to account for both this 'robust averaging' phenomenon and prolonged RTs for more variable arrays if decision values are 'squashed', for example by being passed through a soft-threshold nonlinearity (i.e. a sigmoid function) before being averaged and integrated across time (de Gardelle & Summerfield, 2011). One computationally parsimonious manner of implementing this transformation is to recode the feature value of each element according to the logarithm of the probability ratio (LPR) between the two response options, given the history of feedback for this stimulus value. This transformation mirrors what might occur in a simple, biologically plausible neural network in which inputs are mapped onto binary responses via weights updated with a supervised, winner-takes-all rule (Ratcliff, Van Zandt, & McKoon, 1999).

For each element x, we thus defined the LPR value of x (see eq 1 and methods) which expresses the probabilistic evidence conveyed by the stimulus x in favour of the rightward option (for positive LPR values) or the leftward option (for negative values). In the context of our experiment, this LPR transformation had a sigmoidal shape, by which extreme elements do not 'pull their weight' compared to those near the centre of feature space (where the function is roughly linear). Of note, the likelihood function is sigmoidal in our experiment because of the mixture of Gaussians (4 levels of mean x 3 levels of variance) from which samples were drawn across trials. Consequently, using the average LPR to drive a drift-diffusion decision process could account for the slowing down of response latencies on more variable arrays, because the weight of outlying evidence is muted, reducing the overall input to the decision process on those trials (Figure 1c, lines). In addition, this LPR transformation could capture the downweighting of elements with extreme feature values exhibited in participants' behaviour (Figure 3.1c, right panel).

In what follows, we assess how brain responses of participants can be predicted by the intrinsic uncertainty of each stimulus array in the categorization task. To do so, we defined the quantities U_{Mr} and U_{Vr} to express the effects of manipulating the stimulus mean

and variance in terms of the ‘decision-space’, that is, the space of the LPR-transformed values (see eq 2-5 and methods).

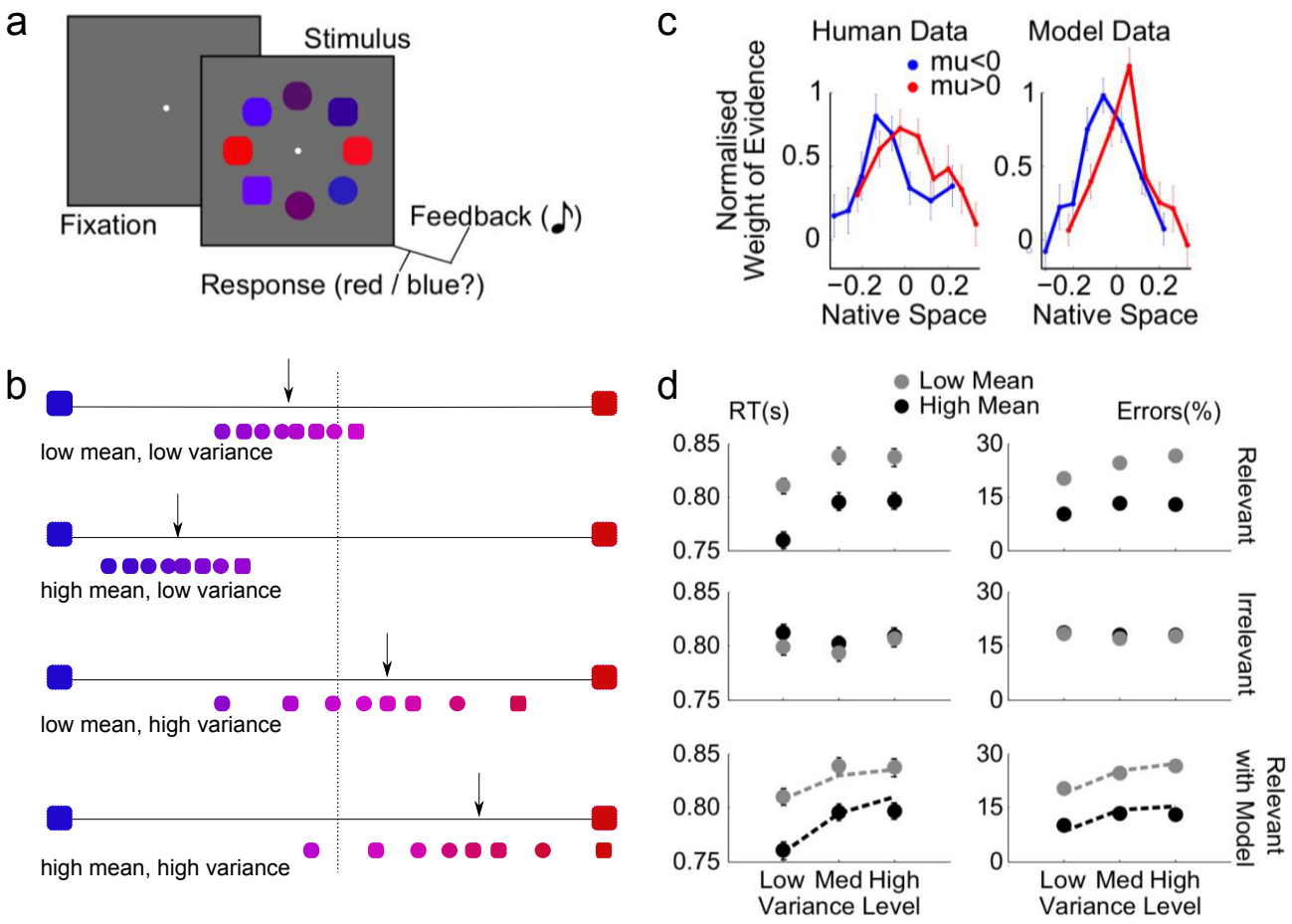


Fig 3.1. Task, behaviour and modelling. (**a**) Schematic representation of a trial: a white central fixation point, was followed after 500ms by the stimulus array, which participants categorised based on either the average shape (square vs. circle) or average colour (red vs. blue) across all elements, with auditory feedback. (**b**) Schematic representation of the possible distribution of the eight trial elements along the colour dimension (Red-Blue). (**c**) Coefficients from a logistic regression in which decision-relevant values, *ranked* in each trial, predicted observers' choices, separately for blue/square (blue) and red/circle (red) stimulus arrays. Higher decision weights appear for inlying vs. outlying elements. The abscissa indicates the average decision value of the elements in each rank, in native space. (**d**) Response times (RTs, left panels) and error rates (right panels), as a function of the mean and variance manipulations. Low mean and high variance correspond to high uncertainty. Top row: effect of the task-relevant manipulations. Middle row: effect of the irrelevant dimension. Bottom row: best-fitting model data (dashed lines) overlaid on the human data for the task-relevant manipulations. The model was fitted to errors only, and RTs are predictions.

3.2.3. BOLD responses associated with mean evidence.

In the first set of analyses, we thus searched for voxels where the BOLD signal correlated positively with the quantity U_{Mr} which reflects the uncertainty due to the mean of the evidence on the task-relevant dimension. In what follows, we report uncorrected statistics but all reported activations exceeded a cluster-corrected threshold of $p < 0.05$ unless explicitly noted in the text.

The results are shown in Figure 3.2 and 3.3 (red blobs) and reported in Table A.1.1 U_{Mr} was positively associated with the BOLD signal in the dorsomedial prefrontal cortex (dmPFC; Fig 3.3a), [peak: 0, 22, 48; $t_{(20)}$=5.22, p<0.0001] and anterior insular cortex (AINS; Fig 3.3b) [left peak -30, 20, 2; $t_{(20)}$ = 6.07, p<0.00001; right peak: 34, 24, -2; $t_{(20)}$ = 6.2, p<0.00001;]. Note once again that this positive correlation signals higher BOLD signals as LPR approaches zero and responses come into conflict. A positive association with U_{Mr} was also observed in the inferior parietal lobule (IPL) [left peak: -34, -48, 50; $t_{(20)}$=6.22, p<0.00001; right peak: 38, -44, 54; $t_{(20)}$=4.40 p <0.001; Figure 3.v2a, far left] and dorsolateral prefrontal cortex (dlPFC) both anteriorly in Brodmann's area 46 [left peak: -50, 32, 30, $t_{(20)}$ = 4.40, p < 0.0002; right peak: 42, 44, 26, $t_{(20)}$ = 4.76, p < 0.0001) and more caudally in Brodmann's area 8 [left peak: -46, 4, 26 $t_{(20)}$=5.77, p<0.0001; right peak: 50, 8, 26; $t_{(20)}$=4.90, p<0.0001; Figure 3.2a, centre left]. Negative correlations with U_{Mr} (i.e. higher BOLD with growing unsigned LPR) were observed in the ventromedial prefrontal cortex and posterior cingulate (Fig A1.2).

3.2.4. BOLD responses associated with evidence variability. Our next step was thus to assess whole-brain responses to the variability of the evidence. As for the results reported above, all results shown remained significant following correction for multiple comparisons at the cluster level (see Table A.1.2). Here, we report brain regions that were sensitive, across trials, to the standard deviation of the LPR over elements; i.e. U_{Vr} (uncertainty due to variability) as defined above.

Positive correlations with evidence variability (U_{Vr}) were observed in the visual cortex (Figure 3.2d & e, yellow blobs). These reached maxima in separate sites in the middle

occipital gyrus (sVis; visual area 3) which were positively activated when the evidence was more variable [left peak: -30, -84, 14 $t_{(20)}$ =8.15 p<0.000001; right peak: 34, -76, 14, $t_{(20)}$ = 7.37, p<0.00001; centre right panel] and at a more ventral site overlapping with the inferior occipital gyrus (iVis) [$t_{(20)}$=4.02, p<0.0004; far right panel]. We also observed strong positive correlations with evidence variability at sites coextensive with the superior parietal lobule (SPL) [left peak: -18, -60, 54, $t_{(20)}$=5.56, p<0.0001; right peak: 26, -56, 54, $t_{(20)}$ = 5.90, p<0.00001; middle panel].

As shown in Figure 3.3a (green blobs), however, *negative* correlations with U_{Vr} were observed in the dorsomedial prefrontal cortex (6, 32, 42, $t_{(20)}$ = 5.73, p<0.00001;). We also observed negative correlations with evidence variability in the anterior insula, with symmetric peaks in the left (-30, 24, -2, $t_{(20)}$ = 3.95, p < 0.001) and right (26, 20, 2, $t_{(20)}$ = 3.95, p < 0.001) hemispheres, although these fell just short of corrected statistical thresholds (Figure 3.3a, right). These negative correlations with U_{Vr} denote voxels where the BOLD response increased as arrays became more homogenous.

3.2.5. BOLD response to the mean and variance of the irrelevant dimension. Our stimulus array consisted of a relevant and an irrelevant dimension (either shape or colour). This afforded us the opportunity to carry out precisely parallel analyses on the dimension of the array which was irrelevant to the decision. One small cluster was found to negatively correlate with U_{Mi} ($t_{(20)}$ = 4.23, p<0.001), but it did not survive correction for multiple comparisons (Table A.1.1.3). No activations were observed in any regions correlating with the variance of the 'evidence' on the irrelevant dimension (U_{Vi}), although positive correlations with the irrelevant variance were observed in the visual cortex at very lenient thresholds (p < 0.005 uncorrected). The lack of reliable correlation with the statistics of the irrelevant dimension confirms that the above-described effects are due to processing of decision-relevant signals.

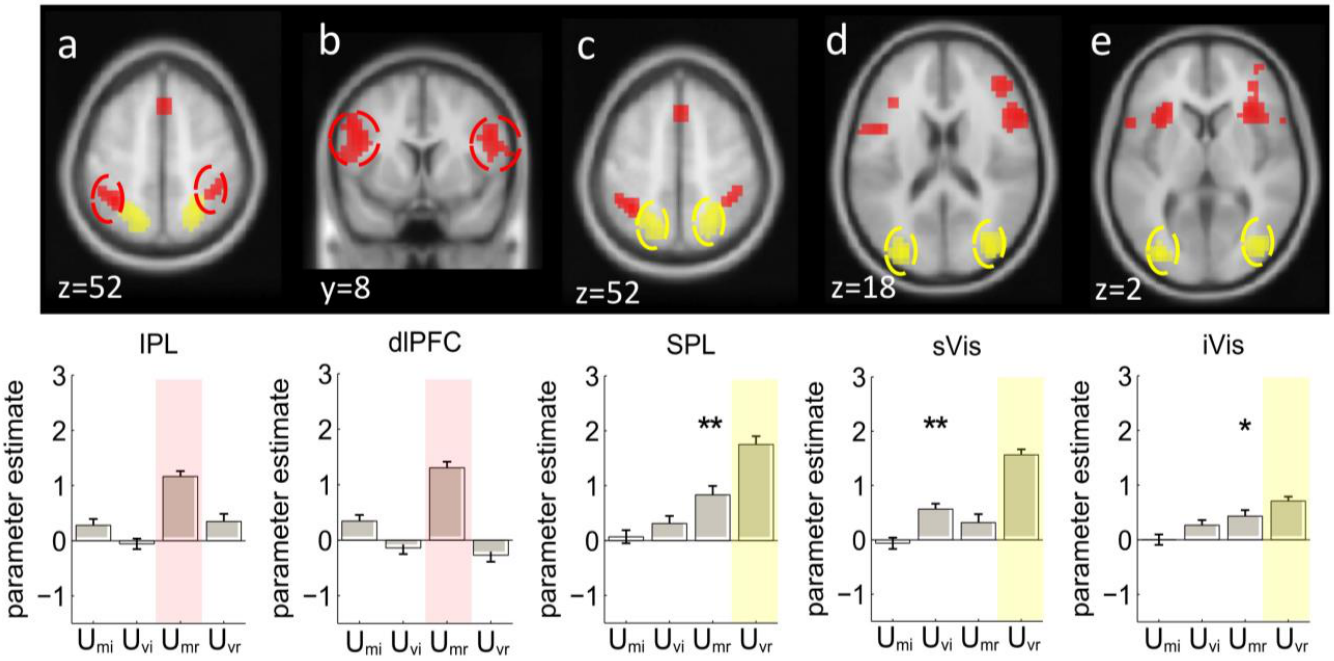


Figure 3.2. Imaging results from dorsolateral prefrontal, parietal and visual cortices. Top row: voxels where BOLD activity was responding to positively correlated with mean-related uncertainty (U_{Mr}, in red), and positively correlated with variance-related uncertainty (U_{Vr} , in yellow). All activations are rendered on the template brain of the Montreal Neurological Institute with an uncorrected threshold of $p<0.001$ (see text and tables for full list of activation and peak coordinates). Bottom row: average parameter estimates from a 5 mm sphere centered on the peak activation from the cluster highlighted with a dashed ellipse, for regressors encoding uncertainty due to the mean and variance for relevant and irrelevant dimensions. Stars indicate significance: *$p<0.05$, **$p<0.01$, ***$p<0.001$. Red and yellow shading denotes the condition used to define the ROI. **(a)** Voxels in the inferior parietal cortex (IPL) responding to U_{Mr} (**b**) Voxels in the dorsolateral prefrontal cortex (dlPFC) responding to U_{Mr} (**c**) superior parietal cortex (SPL) showed a positive correlation with U_{vr} (and U_{Mr} – see bar plot). (**d,e**) We subdivided a large region of visual cortex showing activity positively correlated with U_{vr} into superior (sVis) and inferior (iVis) regions.

3.2.6. Overlapping and nonoverlapping responses to evidence mean and variability.

To compare responses to mean and variance, we extracted regions of interest focussed on peaks responding positively to U_{Mr} and tested their sensitivity to evidence variability (i.e. U_{Vr}), and vice versa (Figure 3.2b). ROIs were defined as a 5mm radius sphere centred on the peak activated voxel in each cluster. Of note, these analyses were strictly independent from one another, as the mean and standard deviation of the feature values in the array were orthogonal by design, and remained so after conversion to LPR values (mean r = -0.01, p = 0.29, t-test of Fisher's z-scores against zero). For completeness we additionally verified sensitivity to the irrelevant mean and variance via this approach, making separate plots for U_{Vr} and U_{Mr} (correlations with the task-relevant variables) and U_{Vi} and U_{Mi} (correlations with the task-irrelevant variables; see also Figure 3.2b).

Voxels responsive to U_{Mr} in the IPL and dlPFC failed to respond to evidence variability (all p-values > 0.05; Figure 3.2b far and centre left panels). However, superior parietal lobule regions sensitive to U_{Vr} were additionally responsive to the U_{Mr}, i.e. to the proximity of the mean LPR to zero, in both hemispheres [left peak, $t_{(20)}$=2.334, p=0.02; right peak, $t_{(20)}$ = 2.732, p=0.007] (Figure 2c). Alone among these ROIs the more superior visual region responded additionally to the task-irrelevant variance U_{Vi} (Figure 3.2d), whereas the more ventral visual region responding to variance also responded to U_{Mr} [$t_{(20)}$=1.80, p<0.04] (Figure 3.2e).

By contrast however, peak voxels in the dorsomedial PFC and anterior insula identified by virtue of their response to mean evidence were additionally responsive to evidence variance (Figure 3.3). Negative correlations with U_{Vr} were observed in the dmPFC ($t_{(20)}$=5.16, p<0.00003; Figure 3.3a) but also reliable in the right ($t_{(20)}$=2.44, p<0.012) and left ($t_{(20)}$=2.88, p<0.005) AINS (Figure 3.3b). Note once again that negative correlations with U_{Vr} signal increasing BOLD signal as the array becomes more homogenous.

These results are striking for two reasons. Firstly, they show a sharp dissociation between two portions of a network frequently activated by decision uncertainty or the demand of action selection. Parietal sites responded to the level of relative evidence in the stimulus array – quantified as the proximity of the LPR to zero – but either failed to respond (at

inferior sites) or responded positively (at superior sites) to the variability of the evidence. Medial prefrontal and insular sites, by contrast, were equally responsive to the LPR, but responded *negatively* to the variability of the evidence. In other words, more positive-going BOLD signals were observed in the dmPFC/AINS when the stimulus array was more homogenous. This is particularly surprising, because the dmPFC in particular has been proposed to be sensitive to the likelihood of an error (J. W. Brown & Braver, 2005) or to time-on-task (Grinband et al., 2011), whereas here participants were both faster and more accurate on trials with more homogenous feature values. We thus sought to validate these findings with a further series of control analyses.

3.2.7. Correlations with mean and variance of raw feature values.

Could the negative correlation between dmPFC and AINS BOLD signal and evidence variability be due to some artefact of our log-probability transform of decision values? To rule out this possibility, we conducted the same analyses as described above but using the statistics of raw ('native space') feature values $|\mu|$ and σ rather than their LPR-transformed counterparts (see Table A.1.2). Globally, the results were qualitatively similar, but statistically more modest. In Figure 3.3c, we show the overlapping clusters responding positively to mean-related uncertainty (negative correlation with $|\mu|$) and negatively to variance-related uncertainty (positive correlation with σ) in the medial PFC for decision-space and native-space analyses. Similar results were obtained for the AINS and visual and parietal cortices.

To additionally ensure that our unexpected findings were not due to misfitting of our basis function (canonical haemodynamic response) to the data, we reanalysed our whole-brain data using a finite impulse response (FIR) filter (which makes no assumptions about the shape of the BOLD response) and plotted the HRFs for low and high mean $|\mu|$(Figure 3.3d, left) , and low, medium and high standard deviation σ (Figure 3.3d, right) separately (i.e. in native feature space). The peak BOLD response averaged across 4s and 6s post-stimulus onset confirmed the pattern of previous analyses, with larger responses to smaller values of $|\mu|$ ($t_{(20)} = 2.414$, $p<0.0127$) and larger responses on trials with low values of σ ($t_{(20)}=2.390$, $p<0.014$). There was no interaction between mean and variance observed.

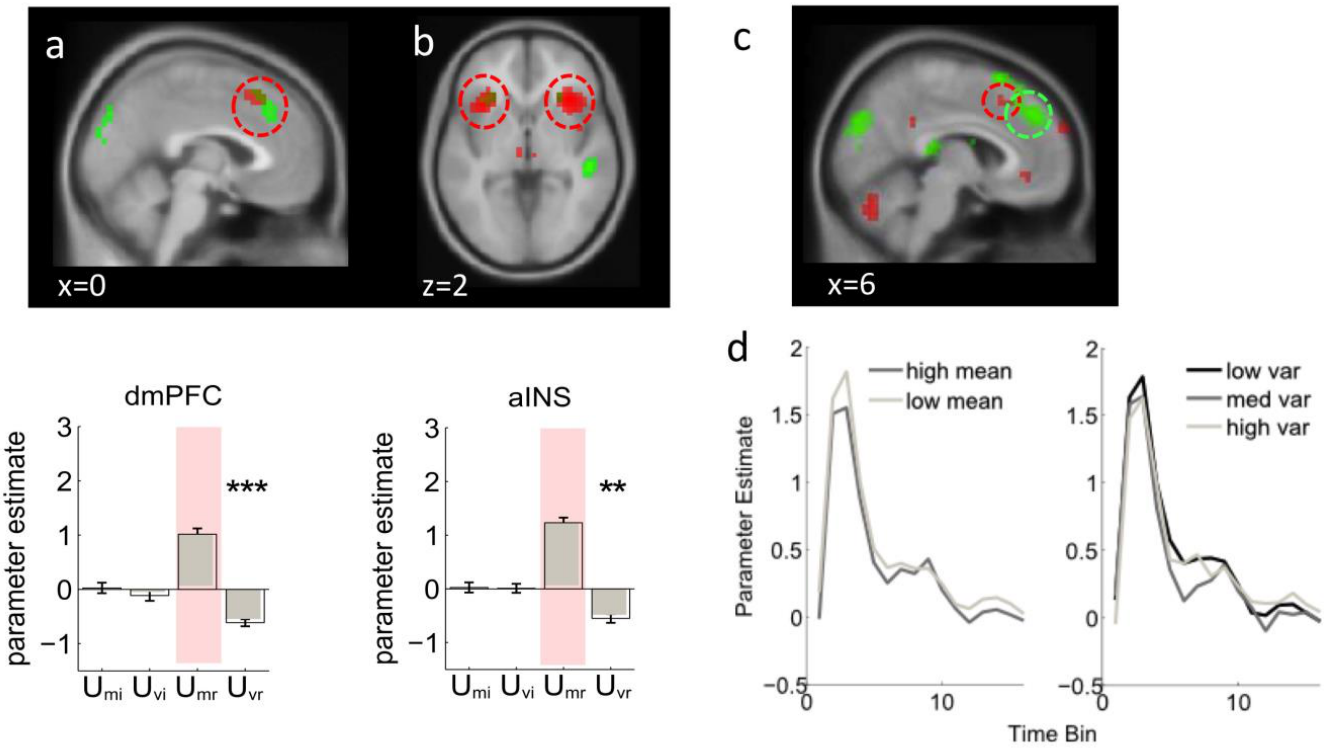


Figure 3.3. Imaging results from dorsomedial prefrontal cortex and anterior insula. (**a**) Upper panel: voxels responding positively to uncertainty due to the mean (U_{Mr}; red) and negatively to uncertainty due to the variance (U_{Mr}; green) rendered onto a saggital slice of the MNI template brain. The corresponding bar plot shows mean responses extracted a sphere of 5 mm radius around the peak voxel for the highlighted cluster, with stars denoting the statistical significance as in figure 2. **(b)** Same results for an axial slice showing the AINS. **(c)** Correlations with native space mean (positive correlation with |μ|, red) and standard deviation (negative correlation with σ, green) in dmPFC, rendered onto a saggital slice at a threshold of $p < 0.005$ uncorrected. The scale indicates the t-value. (**d**) Left panel: haemodynamic response functions (HRFs) generated from a finite impulse response (FIR) model for the dmPFC ROI (5mm sphere extracted from peak of native space activation) for low mean (i.e. |μ| close to category boundary; light grey) and high mean (dark grey). X-axis shows time in scans (2s). Right panel: HRFs for low (black), medium (dark grey), and high (light grey) variance.

3.2.8. Voxelwise correlations between effects of mean and variance.

Our statistical approach involved identifying voxels that responded to evidence mean and testing their sensitivity to evidence variability, and vice versa. One limitation of this approach is that two adjacent but nonoverlapping clusters might become smeared into one by spatial smoothing, potentially giving rise to the spurious impression that a single region responds to both variables. We thus conducted a further analysis in which we correlated the response to U_{Mr} and U_{Vr} in a voxelwise fashion using the unsmoothed data, and converted the correlation coefficients at each voxel to a Fisher's z-score, permitting parametric statistics at the group level. The resulting group statistical maps, which were only smoothed *after* correlations were calculated, indicated voxels where there were significant correlations between the response to U_{Mr} and U_{Vr}. The results are shown in Figure 3.4a and 3.4b. We observed a cluster of negative correlation between response to these two variables in the vicinity of the dmPFC/ACC, with peaks at -2, 28, 30 ($z = 3.17$, $p < 0.001$, uncorrected) and -6, 36, 30 ($z = -3.34$, $p < 0.0005$, uncorrected; Figure 4a, left). Additional clusters were observed at the left (-38, 20, -6, $z = -3.43$, $p < 0.0003$, uncorrected) and right (42, 24, -2, $z = -3.32$, $p < 0.0005$, uncorrected) AINS (Figure 3.4a, right). Within the ROI defining the dmPFC by its sensitivity to U_{Mr}, the average Fisher's z-score was also significant ($z = -2.39$, $p < 0.009$). In other words, those voxels that responded positively to U_{Mr} (higher signal as the mean LPR approached zero) tended to respond negatively to U_{Vr} (higher signal as evidence became more homogenous) and vice versa. No positive or negative correlations between response to U_{Mr} and response to U_{Vr} were observed in other regions sensitive to U_{Mr}, such as the parietal cortex.

3.2.9. Neural correlation with reaction time in dmPFC. One reason why the negative correlation with U_{Vr} in the dmPFC is surprising is that this region has often been reported to exhibit a positive correlation with reaction time (RT), and here RTs are longer when feature variability is high. We thus conducted a further analysis in which a regressor whose height was parametrically modulated by RT on each trial was included. As expected, the dmPFC region (defined as above) correlated positively with RT ($t_{(20)} = 2.38$, $p<0.0139$) as

well as U_{Mr}. Critically however, the inclusion of RT left the negative-going response to variability in the dmPFC region defined by positive correlation with U_{mr} intact (Figure 3.4b and $t_{(20)}=5.27$, $p<0.00002$). In other words, the negative correlation between BOLD activity in the dmPFC and evidence variability persists even though the latter elicits greater behavioural cost.

3.2.10. Correlation with distance-to-bound in the dmPFC.

One explanation for this counterintuitive finding is that information that falls close to the category boundary is processed with enhanced gain. This proposal follows naturally from our modelling approach in which individual stimulus values were sigmoidally transformed and averaged across elements within the array, before they contribute to an "integration-to-bound" decision process (in this decision process, some information corrupted with noise is accumulated over time until a criterion is reached that triggers the response, see e.g. Ratcliff & Mackoon, 2008). One consequence of the sigmoidal transformation is that information closest to the boundary will have the most powerful impact on choices, as demonstrated by the 'robust averaging' behaviour observed in this cohort (and previously) whereby elements falling far from the boundary are downweighted in the choice.

One plausible neural implementation of this model is one in which the gain of encoding of decision information is strongest near to the category boundary. We thus attempted to formalise this idea, calculating a new quantity that indexed total distance of all elements to the category boundary, in decision space. We call this variable distance to bound (*D*).

$$D = \sum_{k=1}^{8} \left| LPR(x_k) \right|$$

D thus indexes the total absolute divergence of the evidence from the category boundary, summed across the whole array. Under the conditions created by our experiment, *D* is very highly correlated with evidence variability (average r = 0.87) and uncorrelated with U_{Mr} (average r = 0.38 across the cohort).

For completeness, we conducted a separate analysis in which we used *D* and U_{Mr} to predict brain activity on a trial-by-trial basis (Figure 3.4c). Because of the strong correlation between *D* and evidence variability, it is not surprising that we observed a robust negative response to *D* in the dmPFC (peak: 10, 28, 42, $t_{(20)}$ = 4.94, p < 0.00004) in a whole-brain analysis. We additionally observed a significant negative response to *D* in the peak dmPFC

voxel responsive to U_{Mr} ($t_{(20)} = 3.71$, $p < 0.001$). In other words, the dmPFC correlates negatively with the absolute distance of all elements to the decision bound, or conversely, correlates positively with proximity-to-bound.

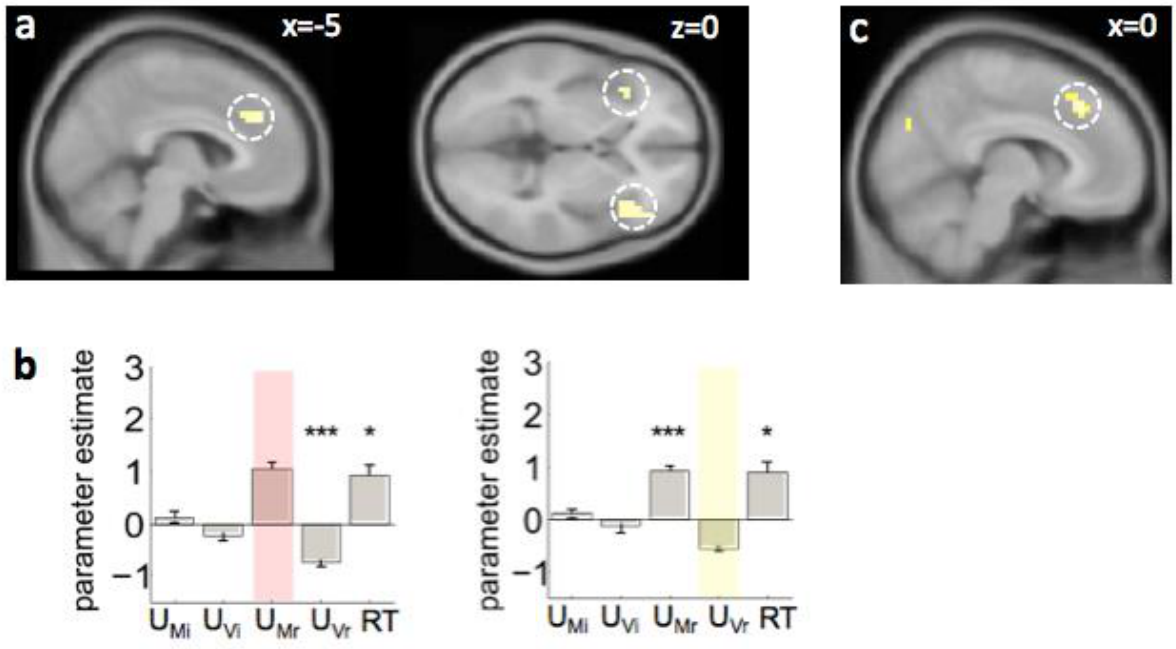


Figure 3.4. Additional imaging results (**a**) voxels where there was a significant negative correlation between the response to U_{Mr} and the response to U_{Vr} shown on a saggital (left panel) and axial (right panel) slice at a threshold of $p < 0.001$ uncorrected. (**b**) results of an anlaysis in which reaction time (RT) was included in the design matrix. Bar graphs show parameter estimates for response to the mean and variance of task-irrelevant (U_{Mi} and U_{Vi}) and task-relevant values (U_{Mi} and U_{Vi}) and reaction time (RT), for a dmPFC region of interest defined by its significant response to U_{Mr} (left panel) and U_{Mv} (right panel). Stars indicate significance: *p<0.05, **p<0.01, ***p<0.001. Note positive correlation with RT in each region, and that all effects described persist even once RT is included. (**c**) voxels showing a negative correlation with D, indexing the absolute distance to bound of all of the elements in the array, rendered on a saggital slice at a threshold of $p < 0.001$ uncorrected.

3.3 Discussion

This study aimed to identify whether the mean and variance of decision relevant evidence rely on independent neural networks to mediate their independent behavioural effects. Correlating changes in these two types of uncertainty with changes in BOLD revealed partially overlapping patterns of activation across a wide range of brain regions.

The results presented above suggest that neural decision processes incorporate at least two sources of perceptual uncertainty across a wide range of brain regions. Although a distinction is often drawn between multiple sources of uncertainty in value-based decision making (O'Neill & Schultz, 2010; Rushworth & Behrens, 2008), it has rarely been addressed in the perceptual domain. Although a large network of brain regions (and potentially neurotransmitters (Yu & Dayan, 2005)) have been shown to be sensitive to perpetual uncertainty, it was unclear whether these activations were driven primarily by changes in evidence variability, or the distance of the average value from the category boundary.

The form this probabilistic transformation takes in the present task means that evidence values lying in the most extreme point in decision space for a given array contribute less to the final decision that would otherwise be expected if all elements received equal weighting. Although equal weighting for all elements is the mathematically optimal solution, for neural systems with limited resources, focusing available resources on the most informative distinctions will be of most use. For a categorisation task, arguably the most important distinctions are those that fall around the category boundary. The LPR model provides just such a weighted space, in that the greatest resolution falls at the category boundary (i.e. the steepest portion of the sigmoid), whilst the more extreme portions of decision space are afforded little resources for differentiation. This kind of coding scheme is biologically plausible, given the limited firing rate dynamics of neuronal systems. The benefit of this transformation is therefore that it provides a principled method for distributing representational resources to where they are needed most.

The overall distribution of brain regions sensitive to either source of decision uncertainty (U_m or U_v) broadly replicated previous findings. Notably, large portions of parietal cortex

were sensitive to both U_m and U_v, consistent with the characterisation of this region as the site of evidence integration for perceptual choice (Fitzgerald et al., 2012; Gold & Shadlen, 2007). This activity was clearly decision relevant, as no similar activation was found relating to the irrelevant dimension. Outside parietal cortex, however, responses to U_m and U_v were more selective. In visual cortices, there were positive correlations with U_v (and even for the irrelevant dimension), but no modulation by U_m. This might suggest that the influence of variability of perceptual choice lies at the level of perceptual representation. It should also be noted that it is not clear whether correlations with U_v reflect the transformed decision space values, since U_v and variance in feature space are highly correlated.

Perhaps the most interesting pattern of activity was that observed in medial prefrontal cortex and bilateral anterior insula. Here, there was a within-region dissociation in the direction of the response to U_m and U_v. There was increased activation in response to trials with evidence close to the category boundary, but it was the low variance conditions that elicited the greatest response from these regions. Prominent current theories of mPFC function struggle to account for this pattern. One idea has been that this region is sensitive to the level of response conflict generated from a stimulus (Botvinick et al., 2001). Response conflict is a quantity that reflects the degree to which the current input provides evidence for competing response option. When there is evidence favouring more than one option, response conflict is high, but where there is evidence only for one option, conflict is low. In the current paradigm therefore, response conflict should be high when there is evidence for both response options. This is most likely to happen for the low mean (all the evidence is close to the category boundary) and for the highly variable conditions (evidence is more likely to fall on either side of the boundary). Therefore, this account would predict positive correlations with both U_m and U_v, rather than the observed positive relationship with Um and negative relationship with U_v. A similar prediction is made by accounts of mPFC function emphasising sensitivity to error likelihood (J. W. Brown & Braver, 2005) or surprise (Alexander & Brown, 2011). One proposition, that this region should scale with RT, was directly ruled out as an explanation for the current results (Grinband et al., 2011).

An alternative explanation must therefore be sought for the current results. One explanation for our finding could be that there is an intermixed population of mean- and variance- sensitive neurons, each responding to the orthogonal task manipulations. Single neurons have been shown to be sensitive to alternative sourced of decision uncertainty in value-based choice (O'Neill & Schultz, 2010, 2014). Single neurons could come to reflect the mean through pooling of sensory information from earlier stages in the visual hierarchy. A similar scheme could operate to make “variance sensitive” neurons; these neurons could show increased activity as the range of input they receive increased. However, no scheme of perceptual representation has reported such neurons and it is perhaps difficult to see what their role would be. Furthermore, this scheme would also not explain why there is negative correlation with U_V in this region. Although this possibility cannot be completely ruled out without investigations at the single unit level, the strong within-voxel correlations in mPFC would support an interpretation in which these neurons are responding to a single quantity.

To understand what kind of quantity this might be, it may be useful to consider what the combination of low mean and low variance may signal about an environment. The critical feature of the task environment is the current location of the category boundary, as to perform the task accurately requires a comparison of the current perceptual information against the observer's best estimate of the boundary. This boundary zone appears to be important for human categorisation, since empirical studies reveal increased sensitivity at the category boundary (Shepard, 1987; Tenenbaum & Griffiths, 2001). According to the LPR model, the transfer function mediating the translation of stimulus space values to decision probabilities is sigmoidally shaped. The steepest portion this sigmoid lies at the centre of this space, centred around the category boundary. A steep transfer function means that discrimination sensitivity should be greatest for this portion of decision space. Non-linear representations can be implemented by gain control mechanisms (Reynolds & Heeger, 2009; Schwartz & Simoncelli, 2001). Therefore, neurons representing the region of space closest to the category boundary may be subject to the greatest level of gain. Overall distance to the category boundary (DCB), in decision space, should therefore be negatively

correlated with neural activity (greatest activity for the smallest distances), which is exactly what was observed in the current data.

Overall, these data suggest that perceptual categorisation judgements take into account both the strength of evidence and its reliability, and that this information is used by a broad range of brain regions. Patterns of selectivity for these two sources of uncertainty are perhaps indicative of the role any particular region plays in the decision process – regions that accumulate evidence may become more active in response to any source of uncertainty whereas a region such as dmPFC may use this summary information to derive a more abstracted signal. One missing piece of information, however, is the time point at which this signal emerges. Prioritising information about the category boundary could reflect a post-decision update signal, or it could reflect modulation of the sensory information in the decision phase itself. Both pre- and post- decisional contributions to task performance have been suggested to account for activity mPFC (Blanchard & Hayden, 2014; Kolling, Behrens, Mars, & Rushworth, 2012; Wunderlich, Rangel, & O'Doherty, 2009), so in the next chapter we will use EEG to identify the temporal location of this mPFC signal.

Chapter 4: Using EEG to Investigate the Temporal Dynamics of Perceptual Choice

This chapter aims to replicate and extent the fMRI findings in the previous chapter, using electroencephalographic (EEG) methods. We will take advantage of the high temporal resolution provided by EEG to try and link two sources of perceptual uncertainty with either pre- or post-decision trial periods.

In previous chapter, fMRI was used to demonstrate how the strength (distance to category boundary) and reliability (variability) of decision-relevant evidence can influence activity across a distributed network of brain regions. These sources of uncertainty associated with the current decision could be relevant to many aspects of the decision process. In addition to the stated task, the demands induced by the changing statistics of the environmental input could influence ongoing attentional or learning processes. However, the limited temporal resolution of fMRI makes it more difficult to identify whether and how the mean and variance affect pre- and post-decisional processing. For example, uncertainty is known to modulate the temporal dynamics of perceptual evidence integration (Huk & Shadlen, 2005; Shadlen & Newsome, 2001), but task demands are also known to modulate neural signals at the time of response and potentially contribute to increased cognitive control for any upcoming stimuli (Yeung, Botvinick, & Cohen, 2004).

We therefore turned to EEG to provide an insight into the within-trial effects of changes in the environmental statistics on category judgements. Typically studied using single-unit techniques (Gold & Shadlen, 2007), recent work has demonstrated the sensitivity of electrophysiological signals to time-resolved decision processes. Although fMRI has been used to identify a network of brain regions that respond to forms of decision uncertainty (Grinband et al., 2006; Heekeren et al., 2008; Huettel et al., 2005), EEG is a method much more suited to identifying a signal reflecting the temporal integration of perceptual evidence, since it can track the process at a fine-grained temporal resolution. It has already been applied to tasks reliant on the sequential integration of discrete pieces of information, all demonstrating the importance of sequence position of evidence in

determining perceptual choice (de Lange, Jensen, & Dehaene, 2010; Wyart, de Gardelle, Scholl, & Summerfield, 2012). Furthermore, MEG has been used to show that the signal in motor cortical regions is predictive of a categorical choice made on the basis of slow accumulation of continuous visual information (Donner et al., 2009). Direct manipulation of the uncertainty in each stimulus, however, has been little studied with EEG.

Several studies have indicated that there are electrophysiological components in the early stages of the decision process that are modulated by stimulus uncertainty. Ratcliff et al (Ratcliff, Philiastides, & Sajda, 2009) argued that the amplitude of a "late" EEG component (~300 ms post stimulus) reflected the quality of available evidence in a simple 2-choice image discrimination task. Quality was manipulated by changing phase-coherence of the images, with the highest phase coherence producing almost 100% classification accuracy. Their conclusion was based on fitting a drift diffusion model to behavioural data, and looking for the parameter which co-varied with the amplitude of this component on a trial-by-trial basis. The parameter that was most highly correlated with this component was the drift rate, which here was interpreted to match evidence quality. However, the late component did not show the same relationship with drift rate or evidence quality with a second stimulus set, suggesting that it is not equivalent to the kind of domain-general accumulation mechanism observed in neural populations of the parietal cortex (Shadlen & Newsome, 2001) and prefrontal regions (Ding & Gold, 2012; Heekeren et al., 2006).

O'Connell et al (Kelly & O'Connell et al., 2012) have laid out a set of criteria for the identification of such a domain general accumulation signal. They argued that such a signal should reflect an integration of evidence, rather than moment-to-moment evidence quality, should reach a reliable threshold that precipitates a response and should also be independent of the evidence modality or motor demand. Using a continuous motion integration task designed to avoid the transient responses induced by abrupt changes in visual stimulation, O'Connell et al reported a centro-parietal positivity (CPP) that fulfilled all these criteria – a component that arguably therefore can be used to index evolution of the decision process in human categorisation tasks.

However, as demonstrated in the previous chapter, perceptual categorisation processes are influenced by both strength of the evidence (mean distance from the category boundary) and its variability. Therefore, it should be expected that these two sources of uncertainty would have dissociable influences on the EEG signal during the decision phase. Both O'Connell et al and Ratcliff et al manipulated forms of stimulus coherence to alter the evidence quality and it is therefore unknown to what extent the observed activity is driven by the mean of the evidence or the current level of variability.

Based on the previously presented fMRI data, it seems that the representation of mean and variance interact in several ways. There are brain regions that respond positively to both decreasing strength and increasing variability of the decision-relevant evidence, a pattern of interaction suggestive of a general sensitivity to evidence uncertainty. More selective responses were also observed. Perhaps the most interesting pattern of response, however, was that observed in medial prefrontal cortex. Here, we observed increasing activity for decreasing strength, but decreasing activity for increased variability of the evidence. One outstanding question is whether this activity relates to the pre-or post-decision period. Given the association of this brain region with adaptive control (Botvinick et al., 2001; Carter & van Veen, 2007; Kerns et al., 2004), it is possible that the observed pattern of activity reflects ongoing control processes that adjust the allocation of attentional resources based on the current task demands. In this case, we might expect to see temporal localisation of this repose profile to the post-decision period. Alternatively, we might see this divergent pattern in the decision phase of the trial. This might suggest that this signal is directly relevant for current task performance, rather than reflecting a response to trial outcome.

This project therefore aimed to further investigate the dissociation between the representation of the strength and reliability of perceptual evidence using EEG, a method with sufficient temporal resolution to resolve pre- and post- decisional contributions of evidence uncertainty to perceptual categorisation.

4.1 Methods

4.1.1. Participants

Twenty-one participants aged between 18 and 30 (12 female, 9 male) were recruited from the University of Oxford. All participants were right handed and had normal or corrected to normal vision, with no history of neurological disorders. Two participants were subsequently excluded due to poor EEG data quality (n=1) or failure to follow task instructions (n=1). All participants were compensated at a rate of £10 per hour for their participation. This study was given ethical approval by a local ethics committee.

4.1.2. Stimuli

Stimuli were created and displayed using PsychToolBox (www.psychtoolbox.org) for MATLAB (Mathworks). On each trial participants viewed an array of eight elements (coloured shapes) circularly arranged (radius ~3° visual arc) around a white central fixation point (5 pixels radius). Elements were equally spaced, equiluminant and covered an equal area on the screen (width = height = 50 pixels for pure circle). Stimuli were presented on an equiluminant grey background. Each element was defined by a shape parameter (S) that determined its position on a continuous transition between a square (S=0) and a circle (S=1), and a colour parameter (C) that determined its position on a continuous transition between blue (C=0) and red (C=1). On each trial, the parameter values for each dimension were drawn independently from a Gaussian distribution with mean μ and standard deviation σ. To ensure equal precision of the mean in all conditions, resampling occurred until the sampled trial μ and σ fell within a tolerance of 0.1% of the desired values.

4.1.3. Design

The task design was identical to that described in the previous chapter, with 4 levels of mean distance to category boundary (for analysis purposes, the sign of the category was

ignored to give 2 levels of mean distance to category boundary, which shall be referred to as mu), and 3 levels of variability (implemented by manipulating the standard deviation, and referred to a sigma in what follows), for both relevant and irrelevant dimensions. The exact values of mean differed between participants and were determined on the basis of a thresholding session just prior to the experimental session. The variance was fixed for all participants to the same values (0.1, 0.2 and 0.3). All factors were orthogonalised, to give a 2x3x2x3 design (relevant mu (high/low) x relevant sigma (high/med/low) x irrelevant mu (high/low) x irrelevant sigma (high/med/low)).

4.1.4. Task and Procedure

Participants were individually thresholded using an adaptive procedure prior to the experimental session to determine the mean distance from category boundary necessary to obtain a performance level of 75% correct (low mean condition) and 85% correct (high mean condition) for both the colour and shape tasks.

Participants were asked to make a category judgement based on the average feature value across an array of eight coloured shapes (Fig 4.1A), as in the previous chapter. The category judgement could be made on one of two dimensions: colour (red vs blue) or shape (circle vs square). Only a single dimension was relevant for any given block. Participants completed two blocks of 144 trials of each task type, presented in a random order. On each trial, the stimulus array was presented following the appearance of a fixation point for 100 ms. A response was required within 1500ms of stimulus onset, otherwise the trial was recorded as incorrect. Responses were made with the left and right index fingers on a keyboard, with response mappings counterbalanced across participants, for both tasks. Auditory feedback (800Hz tone for correct, 400Hz tone for errors) was given a fixed interval (800 ms) following the end of the response period. The inter-trial interval was jittered uniformly around 1.7s (+/-0.3s).

4.1.5. EEG Acquisition and Preprocessing

A Neuroscan Synamps2 system was used to record the EEG data with Ag/AgCl electrodes from the following locations: FP1, FPz, FP2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP7, CP3, CPz, CP4, TP8, P7, P3, Pz, P4, P8, POz, O1, Oz, O2. Additionally, pairs of ocular electrodes were placed horizontally and vertically over the left eye. A right mastoid electrode was used as a reference, although data were re-referenced offline using the global average signal. Electrode impedances were kept below 50 kΩ. Data were collected at 1kHz and filtered online with a highpass filter at 0.1Hz and a lowpass filter at 200Hz. For all analyses, data were downsampled to 250Hz.

All preprocessing and data analysis was performed in MATLAB, using the EEGLAB toolbox. The continuous data file was epoched according to stimulus onset, taking 1s before and 3s after stimulus onset. The 1s pre-stimulus period was used for baseline correction. All epochs were visually inspected and any epochs of low quality were marked for rejection and not included for analysis (average rejection rate = 18, 3%). Consistently poor channels were interpolated (only necessary in a single case). Both highpass (30Hz) and lowpass (1Hz) filters were applied. Finally, independent components analysis (ICA) was used to remove movement artefacts.

4.1.6. Behavioural Data Analysis

Behavioural data were analyses with ANOVA and t-tests and all statistics are presented using a significance threshold of $p<0.05$.

4.1.7. EEG Analysis: Event Related Potentials

Following preprocessing, we averaged stimulus or feedback locked time courses to obtain event related potentials (ERPs). For stimulus-locked analyses, the signal was averaged across these electrodes for all subjects for low vs high mu , or low vs med vs high sigma, for either the relevant or irrelevant stimulus dimensions. Data were included from a pre-

stimulus period of 400ms, and up to 1000ms post-stimulus onset. Scalp topographies were also plotted for the effects of mu and sigma, The resulting maps are presented unthresholded, and with maps showing where the associated t-values exceed a set statistical threshold (t-values greater than 3, which correspond to a two-tailed p-value less than 0.01).

At feedback, we again compared the average ERP across subjects on error trials vs correct trials in a time window of -400ms to +700ms.

4.1.8. Computational Model

The future values of each array (in their native space, spanning an arbitrary range of 0-1) were transformed into "decision space" values by the application of the LPR model. This decision space is probabilistic - these values indicate the likelihood of a left or right response given the elements under consideration. Briefly, the likelihood of a particular correct response following the presence of any given feature value is calculated for the arrays seen by each participant. A sigmoid function is fitted to the resulting relationship between each point in feature space and the log-probability of left or right response (the "decision space"). This sigmoid is not an a priori assumption of the model but instead arises due to the intermixing of trial arrays with generative distributions of different mean and variance. For each array, we can then take the average probability value (LPR value) calculated from all eight elements in the array, and the associated standard deviation. We refer to the mean of the probability values as U_m and the standard deviation as U_v. For ease of comparison, we reverse the direction of the mean to scale in the same direction as the variability i.e. high values indicate increasing uncertainty.

4.1.9. Single trial EEG analysis: Regression

We also conducted single-trial analyses using the previously defined decision space values, U_m and U_v. A -400/+1000ms window was taken around either the stimulus onset or

feedback onset, and normalised measures of U_m and U_v were entered into a competitive regressions as predictors for the amplitude of the EEG data (the raw data was first lowpass filtered at 8Hz) at each data point, for each subject's correct trials only. The resulting parameter estimates were also lowpass filtered at 5Hz. All statistical tests were performed at a group level across participants. The statistical significance of the resulting parameter estimates was determined by performing a one-sample t-test at each point. Data are considered significant when t-value was greater than 3, corresponding to a two-tailed p-value of less than 0.01, uncorrected for multiple comparisons. As for the ERP analyses, data were included from 3 centro-parietal electrodes electrodes (Cz, CPz and Pz). Analyses were also replicated for parietal (P3, Pz and P4), occipital (O1, Oz and O2) and frontal (F3, Fz and F4) electrode groupings to address the selectivity of the observed effects. Additional regression analyses substituted U_v for a measure of overall distance to boundary (DCB). This was calculated by taking the sum of the absolute LPR values for each array.

4.2 Results

4.2.1. Behaviour

The mean error rate and reaction times (RTs) were calculated for each level of mu and sigma and entered into a 2 x 3 ANOVA. A significance level of 0.05 is used for all statistical reporting.

As reported previously, participants were both slower and less accurate in judging the average category when both the mean distance to category boundary decreased [RTs: $F_{(1,18)} = 111$, $p < 0.001$; Err: $F_{(1,18)} = 66.2$, $p < 0.001$] or the variance of the stimulus values increased [RTs: $F_{(2,30)} = 23.9$, $p < 0.001$; Err: $F_{(2,31)} = 9.92$, $p < 0.001$]. These effects were independent and the interaction between mu and sigma was not significant for RTs [$F_{(2,34)} = 0.892$, $p = 0.414$] or for accuracy [$F_{(2, 36)} = 1.38$, $p = 0.263$]. Therefore, increasing uncertainty due to either mu or sigma negatively impacted on task performance measures (Fig 4.1B-C).

We also included the mu and sigma of the irrelevant dimension alongside the relevant dimension. There was no main effect of the irrelevant mu (irrmu) or sigma (irrsig) on error rates [irrmu: $F_{(1,18)} = 0.07$, $p=0.802$; irrsig: $F_{(2,32)} = 0.51$, $p = 0.511$], although there was a main effect of irrsig on RTs [$F_{(2,32)} = 4.74$, $p < 0.05$], with faster RTs for the lowest level of variance but no effect of irrmu on RTs [$F_{(1,18)} = 1.74$, $p = 0.204$]. In both the case of error rates and response times, the main effects of relevant dimension mu and sigma remained highly significant. For the full table of results, please refer to Tables A.2.1-2. Therefore, the observed behavioural effects were are largely driven by attention to the task-relevant dimension.

We additionally calculated the weighting functions to relate the position of each element within in a given array with its overall influence on choice. In previous data sets (Chapter 3) it has been observed that elements falling towards the tails of the sample distribution predict the eventual decision more weakly than expected, suggesting that the most outlying elements in a trial array are downweighted relative to those falling in the centre. In the present data, we repeated this analysis by performing a logistic regression on

choice, using the feature values as predictors. These values were first ranked so that the most outlying elements were in positions 1, 2, 7 and 8 whilst the inlying elements occupied positions 3-6. A paired t-test between the average parameter estimates for inlying and outlying elements was then used to test for a significant difference in the weights assigned according to within-trial element position.

This analysis procedure found that, averaging across all trials, there was no significant difference between the parameter estimates for inlying vs outlying elements [$t_{(18)} = 0.56$, $p = 0.583$]. However, splitting trials by the three levels of variance showed that this null effect was driven by the similarity of inlying and outlying elements for low variance conditions [$t_{(18)} = 0.72$, $p = 0.480$]. At both medium [$t_{(18)} = 2.31$, $p < 0.05$] and high [$t_{(18)} = 2.29$, $p < 0.05$] variance, the outlying elements were associated with significantly lower parameter estimates (Fig 4.1D). This was also reflected in a significant interaction between variance level and position [$F_{(2,36)} = 3.44$, $p < 0.05$] This pattern of results replicates a previous observation (de Gardelle & Summerfield, 2011) that there is greater downweighting of outlying elements at high levels of variance. There was no significant difference between inlying vs outlying elements for either low or high mean trials [low: $t_{(18)} = 1.27$, $p = 0.218$; high: $t_{(18)} = 0.04$, $p = 0.966$].

The downweighting of outlying elements can be well accounted for by a computational model in which the integrated decision quantity is based on the log probability ratio associated with each point in feature space. The transfer function derived from this process naturally takes the shape of a sigmoid (i.e. the form was not assumed a priori), which results in the downweighting of outlying elements. In the previous fMRI study, we showed that using the transformed “decision space” values can also better account for changes in neural activity. Since we have replicated the behavioural pattern of increased outlier downweighting across increasing levels of variance, we feel that the use of this transformed decision space can also be applied to the current EEG data.

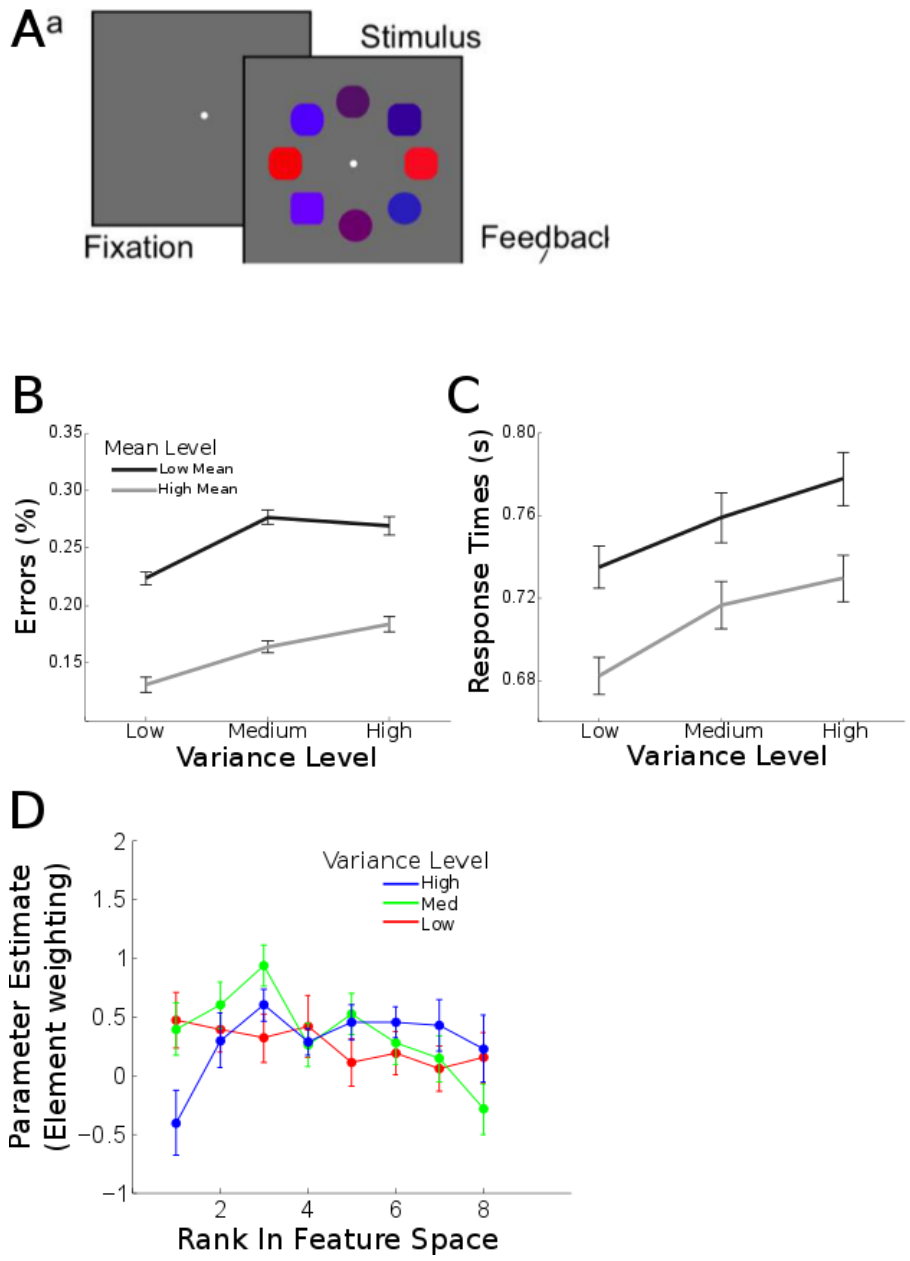


Fig 4.1. Task and Behaviour. A. Participants performed a multi element averaging task in which they were asked to classify an array of coloured shapes on the basis of either the average colour, or the average shape. The relevant feature value was indicated at the beginning of each experimental block. Participants received auditory feedback a fixed interval (800ms) after the end of the response period. B-C Behavioural results. Effects of mu and sigma on error rates (B) and response times (C). X-xis shows the level of sigma (i.e. whether the standard deviation was low, medium or high). The lines show differences for in performance for either the low mean condition (dark line) or high mean condition (light grey line). D. Plotted are the average regression weights associated with the ranked position in feature space for low (red), medium (green) and high (blue) levels of variability. Regression weights are calculated from a logistic regression on choice, using the rank of the elements in feature space from each trial type as a predictor. All error bars show SEM.

4.2.2. ERPs during the Decision Phase

After confirming the independent behavioural effects of mean distance to category boundary and stimulus variability, we performed a parallel analysis on the stimulus-locked EEG data. Reported analyses focus on the central midline electrodes, since previous studies of perceptual choice using EEG have reported that decision-related activity is most prominent at these sites. Overall, the stimulus-locked ERP showed a positive component peaking at approximately 400ms post-stimulus onset, with an earlier and smaller positive deflection at ~100ms (see Fig 4.2A).

We then looked to see whether this response to the stimulus was modulated by the mu or sigma associated with each array. Grouping trials into high, medium or low sigma showed a graded response of the positive component, with the highest amplitude at 400ms for the lowest level of sigma (Fig 4.2B). The scalp topography of the response is plotted in Fig 4.2D, with a statistical threshold (t-threshold of 3, see methods) applied to assess the significance of the difference between levels of sigma (Fig 4.2C). These maps show that there is an early difference (at ~100ms post stimulus) associated with a greater ERP for high than for low sigma. However, the strongest effects emerge between 300 and 500ms post-stimulus - the period for which low sigma was associated with the greatest ERP amplitude. This is unexpected given previous positive relationships between the amplitude of ERP components at this timepjoint (e.g. the P300) and increased decision uncertainty. It is, however, consistent with the negative correlation between BOLD signal and increasing variability we observed in the previous chapter.

In contrast to the effects of variability there was only a small difference between the average ERPs for the two levels of mu (Fig 4.2A), and the thresholded topographies (Fig 4.2C-D) show that this difference is limited both spatially and temporally. There was a very modest difference at approximately 300ms post stimulus at posterior electrodes, corresponding in the ERP plot to a time at which low mu elicited a greater response amplitude than high mu.

To determine whether the observed effects of mu and sigma are due to very basic perceptual processes or to decision-relevant task processing, we repeated the same

analysis for the mu and sigma of the irrelevant dimension (termed irrmu and irrsigma). For irrmu, there was a difference between low and high irrmu conditions in an early post-stimulus component (~200-300ms, Fig 4.3A). The topographies (Fig 4.3C-D) show that there were no differences between high and low irrmu conditions at a later point in the post-stimulus period. ERPs for high, medium and low levels of irrsigma show a qualitatively similar pattern to the the relevant dimension values; there is an early component for which higher levels of irrsigma elicit the greatest amplitude response, and a later component for which the lowest levels of irrsigma are associated with the greatest amplitude (Fig 4.3B). However, the topographies (Fig 4.3C-D) show that only the early (~200ms) difference was statistically reliable, suggesting that this component could be responsible for the weak effect of variability on RTs. Critically, the later component (~400ms) was not associated with a reliable difference between levels of irrsigma. Therefore, these results suggest that the post-stimulus period can be characterised by an early, weaker, response that is driven by several sources of perceptual variability and a later, stronger component that is more reliably associated with decision-relevant information.

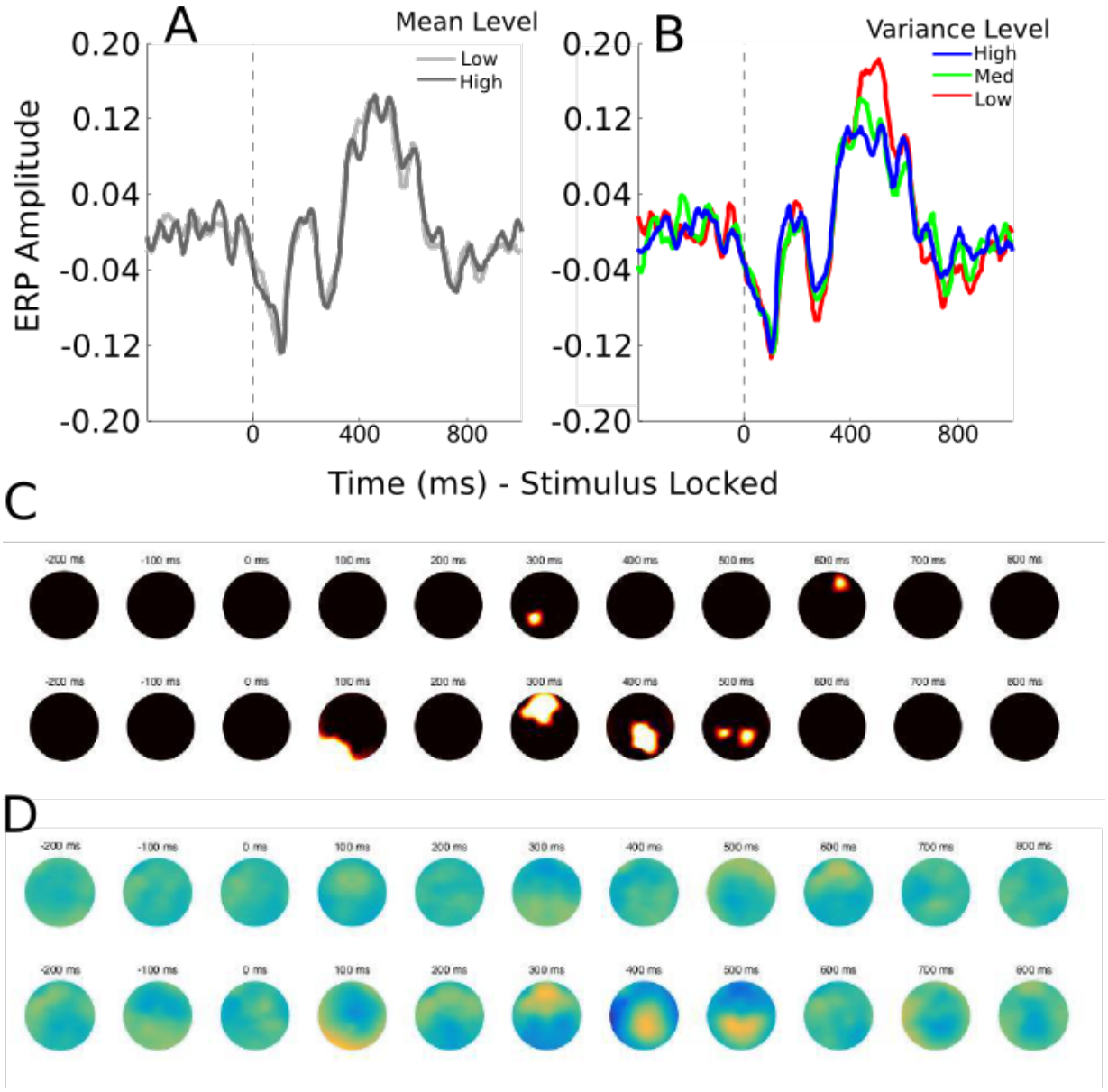


Fig 4.2. ERPs for Relevant Dimension Mean and Variance. A-B. Plots the magnitude of ERPs for the two levels of mu (A; high mean = dark grey, low mean = light grey) and sigma (B; low variance = red, medium variance = green and high variance = blue). Data are shown from 400ms pre stimulus to 1000ms post-stimulus onset (x-axis). Data for ERPs are from central midline electrodes (Cz, CPz & Pz). C-D. Scalp topographies of the effects of mu (top lines) and sigma (bottom lines), calculated from data averaged across 100ms bins, from -200ms to +1000ms for all electrodes. D shows unthresholded amplitude of EEG signal, C shows maps of the associated t-values, thresholded for t-values > 3.

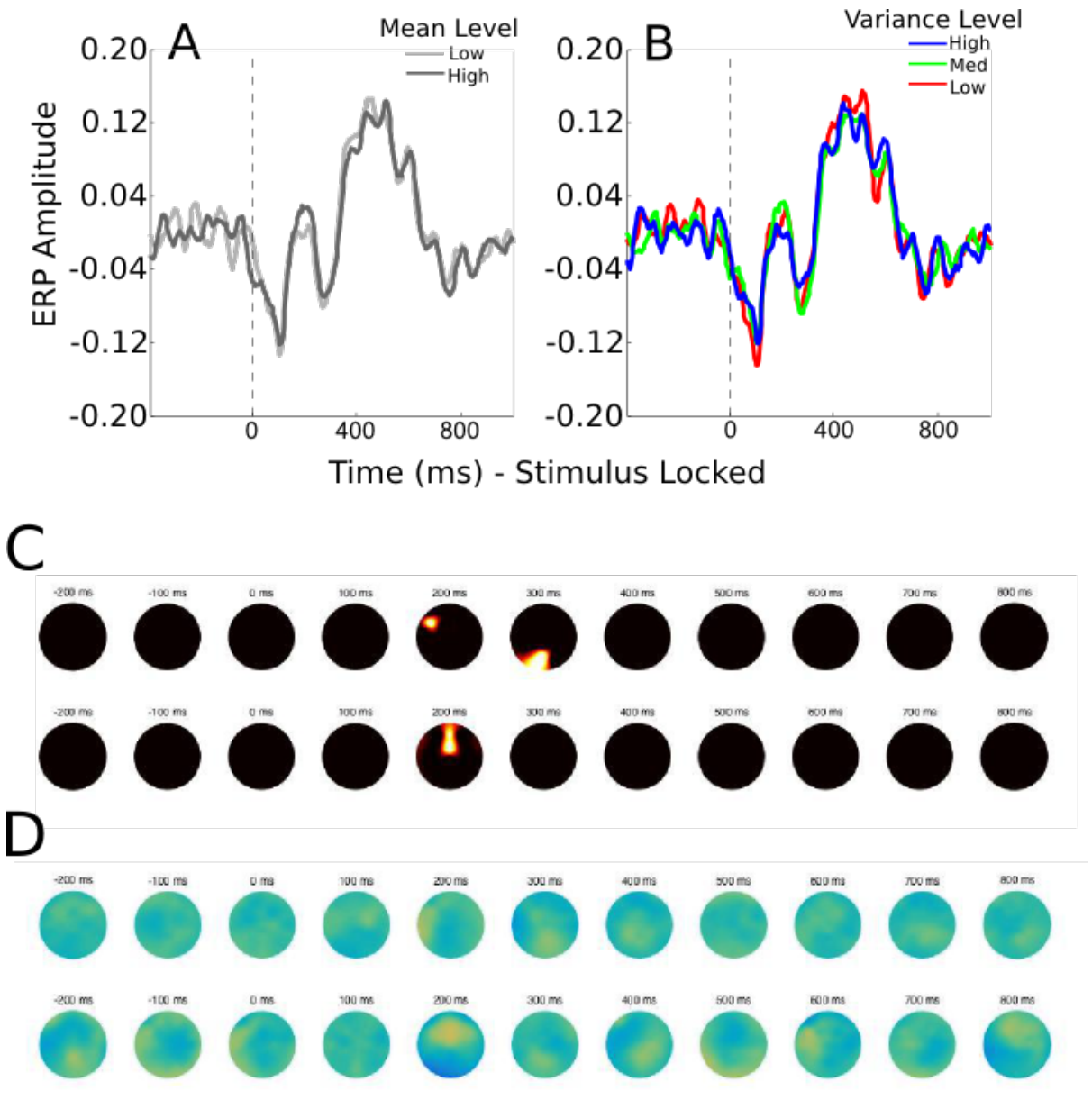


Fig 4.3. ERPs for Irrelevant Dimension Mean and Variance. A-D as for Fig 2, but here levels of mu and sigma are according to the irrelevant dimension.

4.2.3. Regression Analyses – Influence of mean and variability in the decision period.

As the behavioural data here and previously have indicated, choice behaviour can be well described by a model in which decisions are made based on an integration of the probability values associated with each point in feature space. It might then be expected that this transformation of feature values to "decision space" can also account for changes in the EEG.

We therefore calculated two new quantities, U_m and U_v that reflect the mean and variance of the trial decision space values (see methods, also Chapter 3). Both values are positively associated with increasing uncertainty, and therefore larger values correspond to those conditions most likely to elicit slow or incorrect responses. Since these values vary continuously, we can use a regression approach to obtain an estimate of their influence on each trial. To assess their influence on neural activity we entered U_m and U_v as competitive predictors of the amplitude of the EEG for each recorded data point 400ms before and 1000ms post stimulus onset, initially including only data from centro-parietal electodes. Fig 4.4 plots the resulting parameter estimates for U_m (A) and U_v (B).

As can be seen, at approximately 400ms post-stimulus, there is a significant (defined as when t-values exceed 3, equivalent to a one-tailed p-value less than 0.01 uncorrected for multiple comparisons) positive correlation between U_m and signal amplitude [max t-value = 3.56], indicating that there is a greater signal as uncertainty due to the mean increases. This is in agreement with both our own previous findings of a positive correlation between BOLD signal and U_m, and descriptions of a centro-parietal positivity that grows with increasing response times and reflects the increased accumulation time required when the noise associated with the current stimulus is high.

Turning to U_v, however, the opposing direction of correlation is found. Here, peaking at a slightly later time point (~450-500ms post stimulus onset), there is a negative correlation between U_v and EEG amplitude [max t-value = 3.26]. Note that no significant correlations were observed when U_m and U_v were substituted for mu [max t-value = 2.41] and sigma [max t-value 2.22] - the mean and variability in feature space. Therefore, the two measures of decision uncertainty have an opposing influence on the amplitude of the

stimulus-locked, pre-response EEG signal. This replicates the pattern BOLD response to U_m and U_v observed in medial prefrontal cortex in the previous fMRI experiment.

We then repeated this regression approach for three further sets of electrode sites (parietal, prefrontal and occipital, see methods) to investigate the specificity of this response. At the parietal electrode sites, a positive correlation with U_m was also observed [max t-value = 3.10], but no significant negative correlation was found for U_v (although there were strong trends towards both a negative response at ~400ms, max t-value = 2.60, and a later positive correlation, max t-value = 2.94). No significant correlations between amplitude and U_m/U_v were found at either prefrontal or occipital locations.

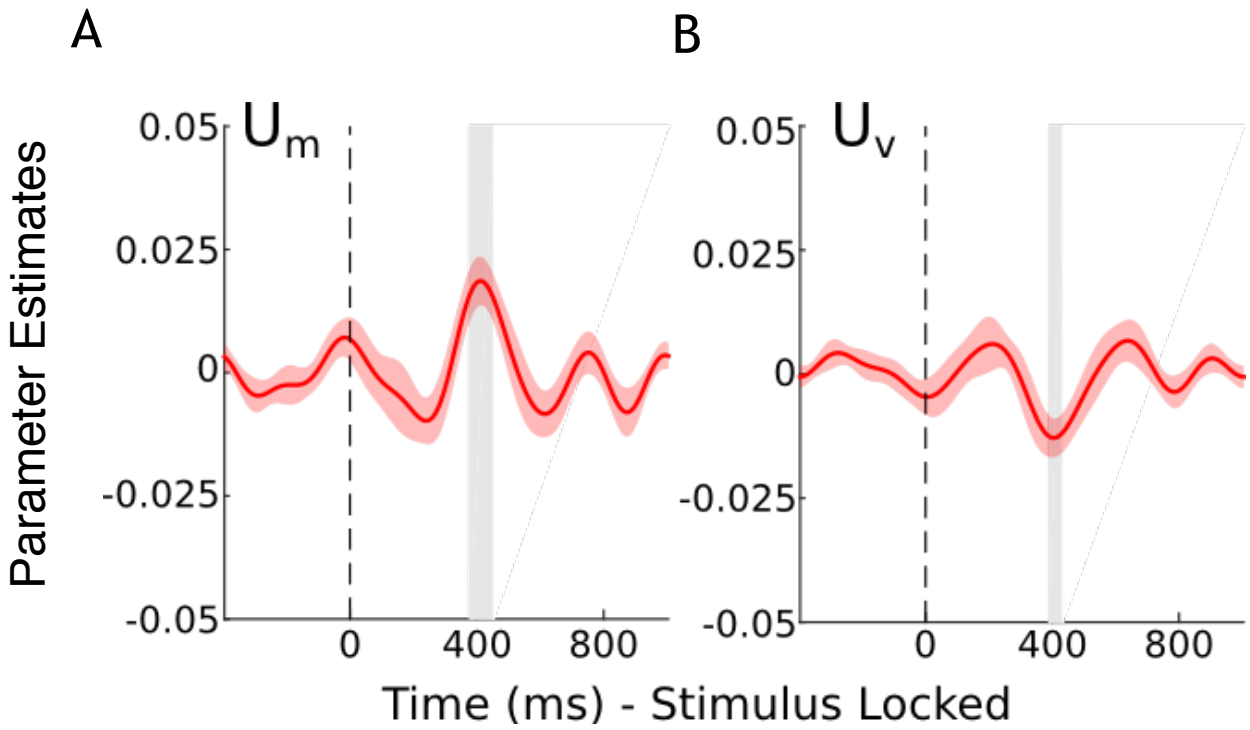


Fig 4.4. Regression Analysis for U_m and U_v in the Decision Phase. A-B Plot the parameter estimates from a regression of U_m (A) and U_m (B) for each data point from -400ms to +1000ms around the onset of the stimulus array. Positive values indicate a positive correlation with increasing uncertainty for both U_m and Uv. Grey sections indicate where a one-sample t-test of parameter estimates across participants indicates a significant (defined as where t>3) deviation from 0.

4.2.4. Alternative Summary Measures: Distance to Boundary.

The previous fMRI data set indicated that the BOLD response of mPFC could be captured by a summary measure indexing the absolute distance to the category boundary of all the elements in a given array, calculated as sum of absolute LPR values for each array. Distance to bound (DCB) is a value that captures the deviation from the central point of decision space, regardless of category and is highly correlated with U_v ($r = 0.8$, across all trials), but not with U_m. Under the LPR scheme, trials on which the values fall close to the category boundary are those with elements that fall in the steepest portion of the transfer function. The increased sensitivity in this portion of feature space may be implemented through increased gain control. As suggested in the previous chapter, this account would predict the greatest activation for trials with the lowest overall distance to category boundary.

We therefore entered DCB into a regression as a predictor of EEG amplitude. The parameter estimates from this regression are plotted in Fig 4.5B. This shows that there was a significant negative relationship between decreasing DCB and EEG amplitude, such that the greatest response was on trials when the mean feature value approached the category boundary. This negative relationship was maintained even when U_m was entered as a competitive regressor alongside DCB. The effect of DCB was limited specifically to the central midline. Whilst a positive effect of U_m was maintained in parietal electrodes, there was no effect of DCB at the same parietal electrode sites. This strengthens the potential that the current positive U_m/negative U_v relationship reflects the same medial prefrontal cortex signal as observed in the BOLD signal in the previous chapter.

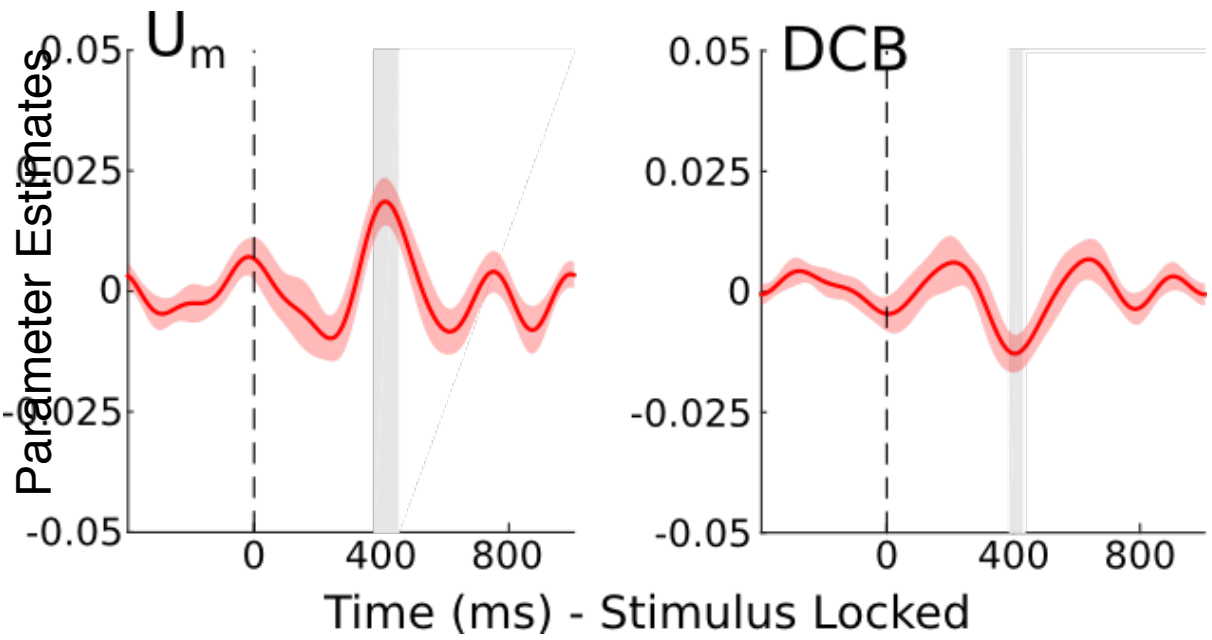


Fig 4.5. Regression Analysis for Um and DCB in the Decision Phase. A-B Plot the parameter estimates from a regression of U_m (A) and DCB (B) for each data point from -400ms to +1000ms around the onset of the stimulus array. Positive values indicate a positive correlation with increasing uncertainty for U_m, and increasing distance from the decision boundary for DCB. Grey sections indicate where a one-sample t-test of parameter estimates across participants indicates a significant (defined as where t>3) deviation from 0.

4.2.5. Regression Analyses: Response to Feedback

In addition to the response period, it was also possible to realign the analysis window to centre on the time of feedback. This phase of the trial contains the most useful information for adjusting future behaviour, since a series of incorrect responses might indicate that the participant's understanding of the category boundary is not complete. Using the same regression approach as for previous analyses, we looked at the modulation of the EEG signal in the period following auditory feedback. Initially, we entered just the main effect of accuracy as a regressor. In this case, a positive parameter estimate would indicate a greater signal magnitude for correct vs error trials and, conversely, a negative parameter estimate would indicate a greater signal magnitude for error vs correct trials. Entering accuracy as a predictor resulted in a profile of parameter estimates that resembles a feedback-related negativity, a confirmation that there was a robust response to the direction of the auditory feedback (Fig 4.6B).

To look at how the response to feedback might be modulated by the stimulus statistics, we entered the interaction between U_m (or, in an independent analysis, U_v) and accuracy into a regression, alongside the main effects of accuracy and U_m or U_v. There was no main effect of U_v, and the interaction between U_v and accuracy was not associated with any modulation of activity in the feedback period. However, a significant positive correlation was found for the interaction of U_m and accuracy (Fig 4.6A), such that the difference between correct and error trials was greatest in magnitude when U_m was lowest. This means that on trials where errors were least common, the amplitude of the EEG in response to an error was the greatest.

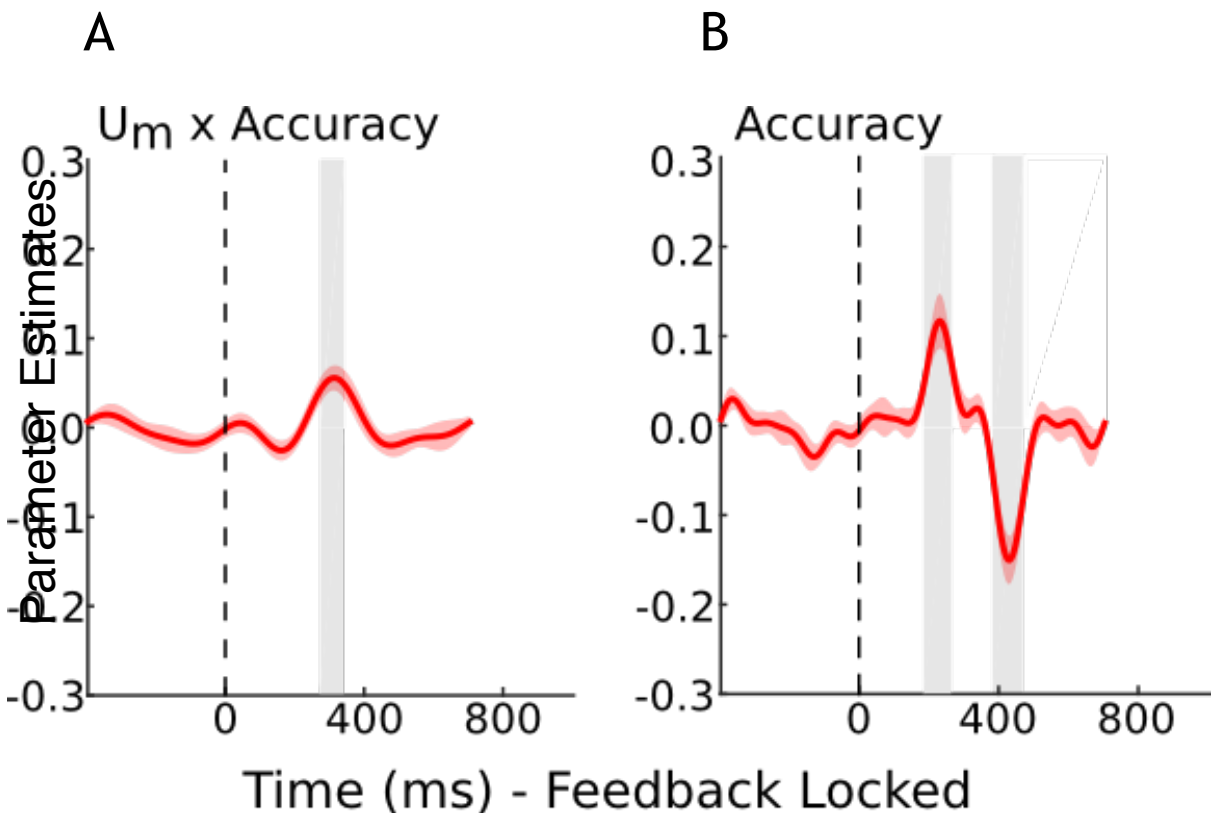


Fig 4.6. Regression Analyses at Feedback. A-B Plot the parameter estimates from a regression of response accuracy (B) and the interaction of response accuracy with U_m, for each data point from -400ms to +700ms around the presentation of auditory feedback. For B, positive correlations indicate where the amplitude of the EEG was greatest for correct trials than for error trials. For A, a positive correlation corresponds to times at which the main effect of accuracy was greater for high vs low U_m. Grey sections indicate where a one-sample t-test of parameter estimates across participants indicates a significant (defined as where t>3) deviation from 0.

4.3 Discussion

The results presented above replicate and extend our previous finding of a dissociation between two sources of perceptual uncertainty – the mean distance from the category boundary and the variance of the trial evidence. EEG allowed us to assess the influence of the mean and variance at both pre- and post- decision points. The strongest effects of decision uncertainty, unsurprisingly, were found in the decision phase. However, uncertainty associated with the mean LPR value also modulated the strength of the feedback signal.

These results replicate several features of the fMRI data presented in the previous chapter. Firstly, they demonstrate that neural activity is well described by a model in which the stimulus values of sensory features on the task-relevant dimension are transformed into probabilistic values, which indicate the most likely generative category given the current evidence. One side effect of this transformation is that the probability values start to saturate as the feature values get closer to the most extreme points in the sensory space (outlying values) (de Gardelle & Summerfield, 2011). This arises from the intermixing of arrays with elements draw from Gaussian distributions with different means and variances. Outlying values are more consistently associated with a single category, and so there is little difference in the associated probability value. For the centre of decision space, however, the association between any given feature value and a response becomes more volatile. The sigmoidal form that this transfer function takes predicts that elements with outlying feature values should have a reduced influence on the categorical choice, a behavioural finding that was partially replicated here.

Secondly, these data replicate the opposing neural correlation with U_m and U_v observed previously in mPFC. In the current data, this pattern of response was unique to central midline electrodes, which is consistent with a mPFC source (Yeung et al., 2004). We were able to localise this signal in time to the decision period in the current study, perhaps suggesting the the mPFC in the fMRI data was also responding during the decision period of the trial. However, the timing of the positive U_m/negative U_v correlation is consistent with reports of a centro-parietal positivity reflecting the integration of evidence to a threshold

- a neural correlate of the decision variable (Kelly & O'Connell, 2013, 2014; O'Connell et al., 2012). An integration to-bound model might, however, predict that there should be a positive relationship between EEG amplitude and U_v, since increasing variability is associated with prolonged response times.

The LPR model, as described in the previous chapter, can provide an explanation for these results. One implication of the sigmoidal form of the transfer function is that the most steepest portion of the function falls at the centre of the space, resulting in the greatest level of sensitivity at the category boundary. As described in the previous chapter, increased sensitivity can be implemented at the neural level by modulating the level of gain applied to neural outflow. A region associated with this gain modulation, therefore, might be expected to show the greatest level of activity when elements tend towards the category boundary. This was shown to be the case for the decision period activity in the present EEG data, again replicating the same profile of activation as in mPFC BOLD signal.

This characterisation of mPFC, as a region controlling the sensitivity of neurons, is not necessarily inconsistent with other theories of mPFC function. According to one interpretation, mPFC is important for learning about value of different courses of action (Rudebeck et al., 2008). In the perceptual domain, perhaps, this could transfer to the control of the influence of evidence at different points in feature space - the evidence provided by some elements will be prioritised over others. Therefore, the “perceptual value” of each element will be reflected in the activity of mPFC neurons.

Although the negative correlation between neural activity and variability was the dominant feature of the current data, there was some indication in the ERPs of an earlier, positive effect of variability at central midline locations. Given the large regions of visual and parietal cortex showing a positive correlation with increasing variability, it was perhaps surprising that a more robust effect was not observed in the EEG data. This is perhaps due to the different sensitivities of EEG and fMRI, perhaps specifically reflecting cancellation in the scalp measure of signals arising from visual cortex anatomy (Vanegas, Blangero, & Kelly, 2013). The early signal was, however, still selective for task-relevant information,

suggesting that the influence of the irrelevant information is dampened from the earliest stages of the trial.

Finally, we observed modulation of a post-decisional, feedback signal in the same central midline electrode by uncertainty. This modulation was, however, specific to U_m and was such that the feedback related negativity was greatest for trials with low Um. One possible interpretation is that this activity reflects a surprise signal (c.f (J. W. Brown & Braver, 2005)). Since participants were more likely to be correct for trials with the lowest U_m, the occurrence of an error would be more unexpected than for trials with high U_m. This explanation is perhaps more aligned with views of mPFC as a control centre, monitoring for expected outcomes to better prepare for the next stimulus (Botvinick et al., 2001; Kerns et al., 2004). If this is the case, then it is interesting that there is no such effect for U_v, since this is also predictive of accuracy. It could be that monitoring activity at the time of feedback is in the frame of response, and therefore is sensitive to information about the identity of the response, but is insensitive to second-order descriptions about the precision of this estimate. This would certainly accord with the evidence suggesting variability primarily influences early perceptual processing.

Together, the present EEG data provide a replication for many of the key features of the fMRI BOLD data observed in the previous chapter and further support the dissociation between two sources of perceptual uncertainty.

Chapter 5: Behavioural Priming Reveals Adaptation to the Summary Variance of Visual Input

This chapter aims to further investigate behavioural dissociations between the mean and variance of perceptual information. Using a priming version of the multi element averaging task, we show that there is a rapid adaptation to the variability of the perceptual evidence. This effect if uniquely driven by the variability, providing a potential functional dissociation between mean and variance in control processes associated with perceptual choice.

The previous experiments demonstrate that categorical judgements can be influenced by the summary statistics of the evidence under consideration. Success in the multielement categorisation task requires accurate computation of the average feature value only, with the additional effect of variance perhaps suggesting that the visual system is tracking the statistical distribution of perceptual evidence. The ability to support the representation of distributions, rather than a collection of individual feature values, allows the observer to make inferences about the likelihood of encountering any feature value in the possible range of inputs. These distributions can also be tracked over time and modulated in response to changes in the external environment.

The benefits of this kind of coding system have been noted for some time (Attneave, 1954; Barlow, 1961; Simoncelli & Olshausen, 2001). This efficient coding hypothesis argued that that matching the internal representation to the statistics of the environment maximises the efficiency of sensory systems. Since the statistics of the environment are constantly changing (e.g. changes in light intensity through one 24 hour day cycle), it should follow that sensory systems are able to incorporate these fluctuations. Indeed, a large body of evidence from invertebrates and primates suggests that adaptation to stimulus statistics is

principle of sensory representation from the retina to higher-order cortical representations (Fairhall, Lewen, Bialek, & de Ruyter Van Steveninck, 2001; Smirnakis, Berry, Warland, Bialek, & Meister, 1997; Wark, Lundstrom, & Fairhall, 2007). However, the links between neural adaptation and behavioural consequence are often not made. For example, recent work in humans and ferrets (Dahmen, Keating, Nodal, Schulz, & King, 2010) showed that adaptation to the mean of the input statistics leads to a bias in discrimination performance, whilst adaptation to increasing variability results in reduced sensitivity.

In humans, there is also evidence demonstrating sensitivity to the sequence of events or stimuli within a fairly short time span. Ide et al (Ide, Shenoy, Yu, & Li, 2013) reported that variability in response times in a stop-signal task could be explained well by a model that assumed participants were tracking the relative probability of either the “stop” or “go” cue appearing. Their model took into account the recent stimulus history when making these calculations, so, following a period including relatively more “go” trials, participants showed faster than average response times. Conversely, a series of “stop” trials would slow response times. This therefore suggests that humans are able to adapt their response strategy to match the local statistics of the environment.

This kind of “micro-adjustment” can therefore represent an adjustment in behaviour that reflects a change in some property of the environment (e.g. trial type or outcome). These adjustments are small, reversible and often rapid changes in performance that promote flexibility, rather than inducing a long-lasting and directional change over time, as might be expected from longer-term learning processes. However, adaptation does not always bring a benefit for behavioural performance, and is perhaps better characterised as shifting the source of costs and benefits to suit the current sensory input regime. Gepshtein et al (Gepshtein, Lesmes, & Albright, 2013) suggest that the mixed findings of costs and benefits in adaptation paradigms reflects an adaptation at the level of the whole stimulus space, rather than an adaptation to the specific stimuli encountered by the observers. Therefore, any adaptation might be expected to both increase and decrease behavioural performance measures, with the benefits targeted to then processing of the most likely sensory inputs.

It might be expected then, that changes in the statistics of the environment will bias this system level adaptation to bring both behavioural costs and benefits. In the psychological literature, it is well known that the response to any given stimulus can be modulated by previous exposure to a similar stimulus. For example, categorisation performance can be enhanced by presenting another category exemplar prior to the target stimulus. This kind of priming phenomenon has been reported for both semantic and perceptual tasks (for reviews see (Schacter & Buckner, 1998; Tulving & Schacter, 1990; Wiggs & Martin, 1998), and seems to reflect a general facilitation for repeated stimuli. Its converse phenomenon, negative priming (Tipper, 1985), reflects slowed RTs to a probe or target stimulus when it requires making a response that was previously required to be inhibited , showing that the congruency of the response associations is also an important factor, beyond simple repetition of visual features. Even from early descriptions, theories of perceptual priming have painted such phenomena as a form implicit memory expressed at the level of the perceptual representation (Tulving & Schacter, 1990). Therefore, priming can potentially be seen as a broad manifestation of perceptual adaptation.

Marchant & de Fockert (Marchant & de Fockert, 2009) compared the effectiveness of modulation by a summary value of the input or by repetition of a single exemplar. They asked observers to make a contrast judgement about a degraded centrally presented circle. This target stimulus was preceded by the brief presentation of a set of eight circles of different size (prime set). Two cases were then considered. In the first, the target circle was equal to the mean size of the prime set. In the second case, the target circle was equal in size to a single member of the prime set. They found that contrast judgements were greater when the target circle matched the mean size of the prime set, an effect that was stronger than the reported contrast enhancement for member repetition trials. Therefore, it appears adaptation can be driven just as easily by a summary representation than by a single exemplar from a given stimulus space. However, it is not clear from this study whether the priming effect was derived from a maintained representation of the full input distribution, or whether the average size acts in the same way as a single exemplar.

In the present study, we aimed to further the results described above and test whether adaptation, in the form of priming, in the multielement categorisation task can be driven by two summary measures – the mean distance from category boundary and the variance.

5.1 Methods

5.1.1. Participants

A total of 94 right-handed volunteers (reporting normal or corrected-normal vision and no history of neurological problems) participated in the three experiments (experiments 1, n = 40; experiment 2, n = 40; experiment 3, n = 14; experiment 4, n = 21). One participant was excluded from Experiment 2 due to data loss in one condition. They provided informed consent and received £10/hour. The study was approved by local ethics committees.

5.1.2. Stimuli

Stimuli were created and displayed using PsychToolBox (www.psychtoolbox.org) for MATLAB (Mathworks) Stimuli were presented on an equiluminant grey background, of a 17 inch LCD monitor with a resolution of 1280 x 1024 pixels, viewed from a distance of approximately 80 cm. Every stimulus was an array of 8 elements. On each trial participants viewed two sequentially presented arrays of eight elements (coloured shapes). In experiments 1 and 2, elements were arranged in a circular formation with a radius of ~3° visual angle around a central fixation point (Fig 1a). In experiment 3, an additional ring with a radius of ~5° visual angle was used. In experiment 4, two prime arrays were presented simultaneously, X cm either side of the central fixation point. Within each array, elements were equally spaced, equiluminant and covered an equal area on the screen. Individual elements were defined by a shape parameter (S) a colour parameter (C) taking values between -1 (most blue/round) and +1 (most red/square) (for full details see supplementary materials). For each array, the 8 values of the S (or C) parameter were drawn from a normal distribution, whose mean μ and variance σ^2 were controlled by design. Feature values were resampled until sample statistics fell within 0.1% of μ and σ^2.

5.1.3. Task

For experiments 1 and 2, which dimension was task-relevant (colour or shape) was counterbalanced across subjects. For experiments 3 and 4, participants judged only colour, which had given slightly more robust effects in the previous experiments. Only one dimension (shape or colour) was the decision relevant dimension within a single session, and participants were instructed to ignore the other dimension.

5.1.4. Thresholding To equalise difficulty across participants and task type (shape or colour) we used an adaptive procedure in which participants made judgments about a target array preceded by a prime array, as in the main task. Each thresholding run was 144 trials long. For experiment 1, the task-relevant mean μ was titrated until performance reached 80% correct on each dimension (shape and colour). For experiment 2, performance was titrated to 75% (μ_1) and 85% (μ_2) for the relevant dimension (either colour or shape). For all experiments, the task-relevant variance σ^2 could take one of two values (0.15, 0.3); in experiment 1, the task-irrelevant variance could also take on these values; in experiments 2, 3 and 4, σ^2 was fixed at 0.15.

5.1.5. Design For experiment 1, we manipulated 5 orthogonal factors: the variance of the prime and the target on the task-relevant and task-irrelevant dimensions, and the congruity of the prime category and the target category on the task-relevant dimension. For experiment 2, we varied orthogonally 4 factors: the absolute level of mean (μ_1 and μ_2) of the prime and target arrays for each category, and the task-relevant variance of prime and target arrays, but fixed the task-irrelevant mean and variance to a single value.

For experiment 3, we fixed the mean to just μ or -μ, but we crossed prime and target array variance with a new factor that encoded the location of the circular prime and target arrays (inner vs. outer ring). The design was thus: target array task-relevant variance (high, low) x prime array task-relevant variance (high, low) x target array location (inner, outer) x prime array location (inner, outer). Experiment 4 included two prime arrays

(presented simultaneously), with one prime array containing a white cue point at its centre, which perfectly predicted the location of the subsequent target array. For this experiment, the irrelevant dimension mean and variance were fixed, along with the relevant dimension mean, as for experiment 3. The target variance was high or low, and the variances of the two prime arrays were orthogonal to each other and to the target (high, low).

5.1.6. Task and Procedure On each trial, a white fixation point (5 pixels radius) appeared in the centre of the screen, followed after 500ms by the prime array. The prime duration was 100ms. The target array appeared after a variable prime-target interval (PTI; 100 ms, 200 ms or 500 ms). For experiment 4, the ISI was fixed to 200ms. Participants had up to 1.5s to respond to the target, using the computer mouse. The target array remained on screen until the response. Response mapping were fully counterbalanced across participants. After a response was made, auditory feedback was given immediately. A high tone (1200 Hz) indicated a correct response whilst a low tone (400 Hz) indicated an incorrect response. Twenty practice trials were given before continuing to the main experiment, which consisted of 768 trials (512 trials for experiment 3 and 1024 trials for experiment 4), with short breaks approximately every 80 trials.

5.1.7. Analysis We log-transformed reaction times for correct trials and analysed the relative influence of the different factors using ANOVAs and multiple regression. An alpha of $p < 0.05$ was used for all statistical reporting.

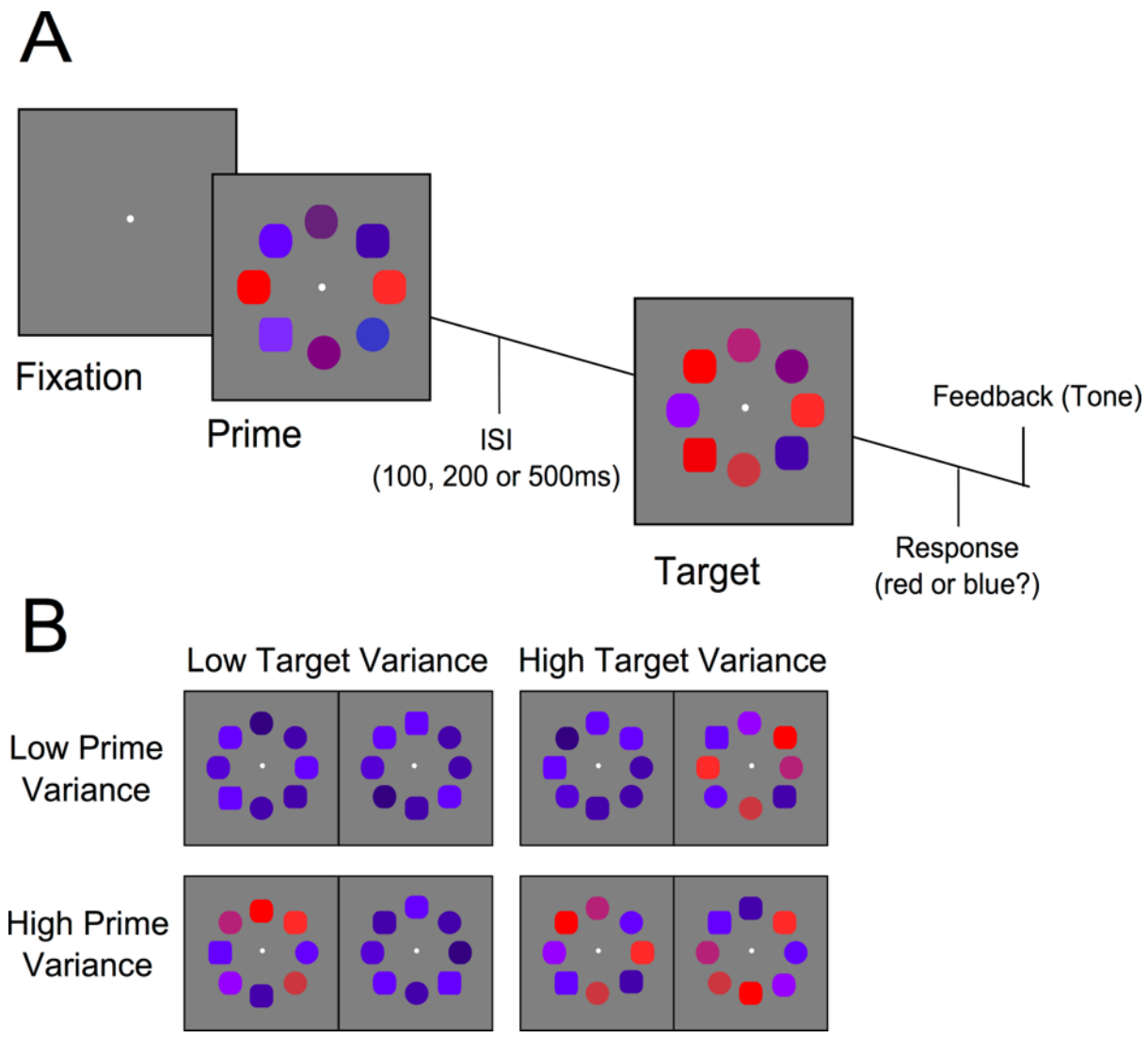


Fig 5.1. Task Description. **(A)** Trial sequence. Observers were asked to classify a target array based on either the average colour (red vs. blue) or shape (square vs. circle). Only one dimension was relevant for the duration of the session. Preceding the target array was an irrelevant prime array, which could appear with a prime-target interval (PTI) of 100, 200 or 500ms. Auditory feedback was given immediately following the response. **(B)** Examples of prime (P) and target (T) arrays with high and low variance on the colour dimensions. Response time facilitation was observed when the variance matched (upper left, lower right panels).

5.2 Results

5.2.1. Experiment 1

In experiment 1 (n = 40), we manipulated the variance (high vs. low) of both task-relevant and task-irrelevant dimensions for both prime and target arrays independently (Fig. 5.1A/B). This allowed us to assess how variability of the prime and target distributions, and (critically) the congruence between their variances, influenced response times (RTs) independently for task-relevant and task-irrelevant feature dimensions. In what follows, all analyses were conducted on log-transformed RTs greater than 300 ms from correct trials only.

As expected, participants responded faster on trials where the prime and target features were drawn from distributions with the same mean value, i.e. from the same category ($F_{(1,39)} = 24.0$, $p < 0.001$). Moreover, increased prime and target variance both slowed response times (prime: $F_{(1,39)} = 5.24$, $p < 0.028$; target: $F_{(1,39)} = 76.9$, $p < 0.001$; see Table A. 3.1.1). Critically however, the congruity of the variance of the prime and target arrays influenced behaviour, with faster responses for trials where the prime and target arrays had the same variance (either high or low), relative to trials where their variance was different (Fig. 5.2a). This was reflected in a significant prime x target variance interaction on RT ($F_{(1,39)} = 27.5$, $p < 0.001$). There was no prime x target interaction on accuracy ($F_{(1,39)} = 0.775$, $p < 0.384$), suggesting that this effect is a facilitation in processing, rather than a due to a change in speed-accuracy tradeoff. The variance of the prime did not impair accuracy, whereas the target variance did ($F_{(1,39)} = 16.0$, $p < 0.001$), as previously reported (see Chapters 3 and 4).

To test whether these findings could be explained by the overlap in individual features between the prime and target arrays, we turned to multiple regression. The prime array variance, target array variance, and their interaction were entered as rival predictors of RT alongside additional regressors encoding the absolute prime-target feature difference for each of the 8 elements in the array (Fig 5.2B), and t-tests were used to assess the deviance of the resulting parameter estimates from zero. Parameter estimates associated

with prime array variance and target array variance were both positive (prime: $t_{(39)} = 4.25$, $p < 0.001$; target: $t_{(39)} = 6.54$, $p < 0.001$), consistent with the previously-described detrimental impact of high variance arrays on decision latencies, whereas those associated with interaction between prime and target variance were negative ($t_{(39)} = 4.79$, $p < 0.001$), consistent with the above mentioned observation that similar variance in the prime and the target facilitated responding. Crucially, these effects persisted even when the eight element-specific differences had been partialled out, indicating that it is a summary statistical representation, not feature information, that is driving priming by the variance.

One alternative explanation not ruled out by these analyses is that RT might be faster when the prime and target offer compatible information about which response to make, but that this response compatibility effect is dampened under high variance. To test this hypothesis, we constructed two further nuisance regressors that encoded (i) the compatibility between prime (P) and target (T), i.e. the absolute of the difference between their task-relevant features ($|\Sigma P_{1\text{-}8} - T_{1\text{-}8}|$), and (ii) the interaction between this quantity and target variance. Repeating the regression analyses described above yielded a weak trend towards a main effect of response compatibility ($t_{(39)} = 1.41$, $p < 0.09$), but no reliable interaction between compatibility and variance ($p < 0.25$). Critically, the interaction between prime and target variance remained strongly significant ($t_{(39)} = 5.19$, $p < 0.001$). Very similar results were obtained using a measure of the absolute activation provided by both prime and target, i.e. $|\Sigma P_{1\text{-}8} + T_{1\text{-}8}|$. These analyses, which are described in Fig A3.1, suggest that our effects are not driven by response compatibility.

To better characterise the processing stage at which priming by variance occurs, we included both task-relevant and irrelevant target variance, prime variance and their interaction in a further analysis. Although weaker than for the task-relevant dimension, the congruity of prime-target variance for the task-irrelevant feature values also facilitated behaviour ($t_{(39)} = 1.72$, $p < 0.05$) consistent with an early, automatic phenomenon (Fig 5.2b). Interestingly, however, there was no crossover variance priming between task-relevant and task-irrelevant dimensions for these trials, i.e. the variance of the task-relevant features for the prime did not interact with the variance of the task-irrelevant features for the target ($t_{(39)} = 0.02$, $p = 0.510$ or vice versa ($t_{(39)} = 0.52$ $p = 0.300$). In other

words, variance priming does not depend on feature-based attention, but does seem to occur within feature-specific channels. Moreover, the existence of independent priming effects for relevant and irrelevant dimensions is hard to reconcile with any explanation based on response compatibility, because no meaningful response mappings were assigned to the irrelevant dimension.

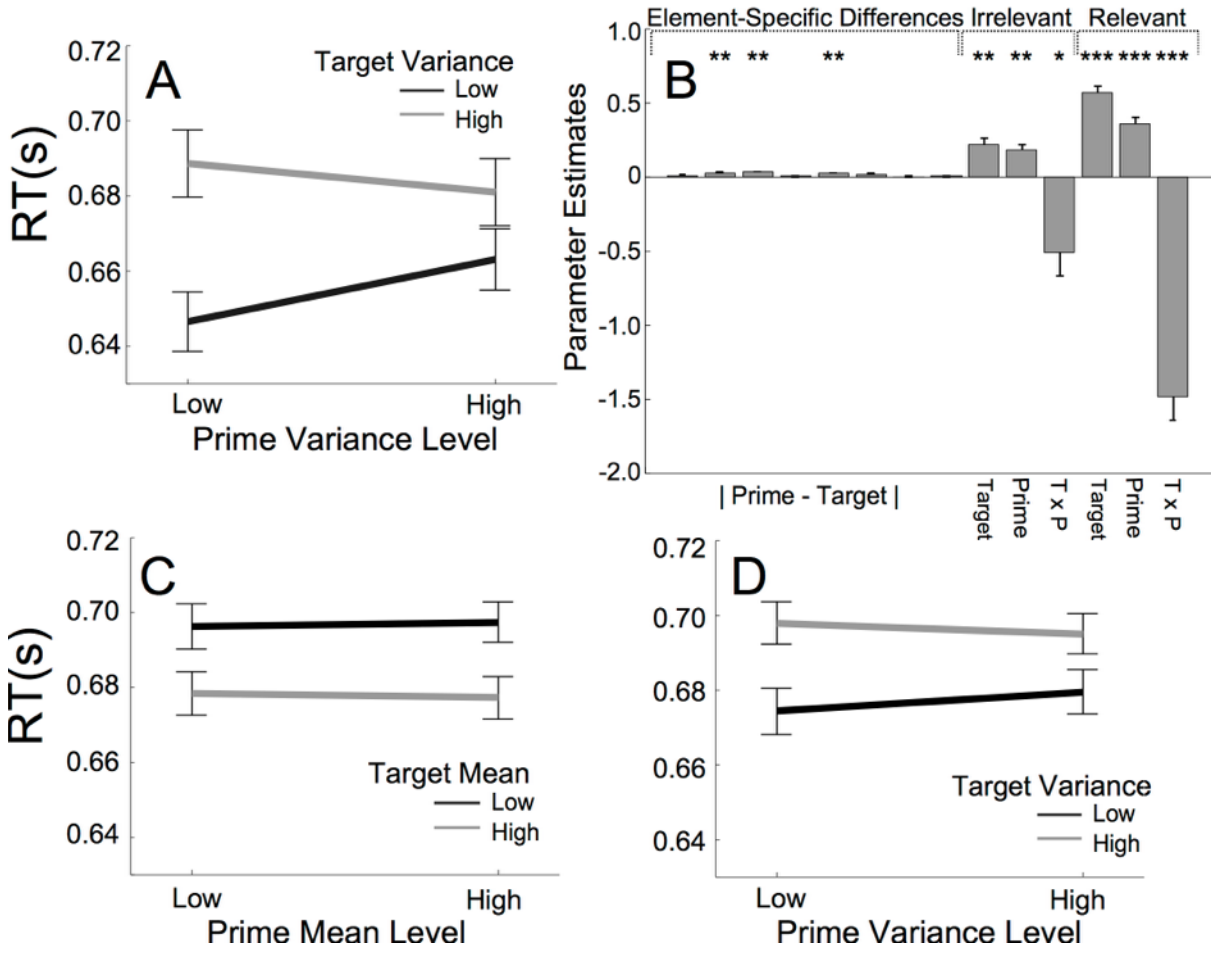


Fig 5.2. Variance Priming. (A) Mean reaction times (RTs) for both levels of prime variance (x-axis; high or low) and both levels of target variance (lines; high vs low). Error bars are standard error of the mean. **(B)** Regression weights are plotted for element-specific differences (|prime-target|), variance of the target ('Target'), prime ('Prime') and their interaction ('T x P') for the irrelevant and the relevant dimension. Stars indicate level of significance as follows: $p < 0.05$ (*), $p < 0.01$ (**) or $p < 0.001$ (***). **(C)** Mean reaction times (RTs) from Experiment 2 for both levels of prime mean distance to category boundary (x-axis) and both levels of target mean distance to category boundary (grey lines). Error bars as for A. **(D)** Mean reaction times (RTs) from Experiment 2 for both levels of prime variance (x-axis) and both levels of target variance (grey lines). Error bars as for A.

5.2.2. Experiment 2.

One outstanding possibility is that facilitatory effects of congruent prime-target array variance could be secondary to nonspecific aspects of the task, such as the relative level of cognitive resources required to judge high and low variance arrays. For example, high variance prime arrays might engage attentional or control processes that confer disproportionate benefit on processing of the high-variance target array, an explanation that has been invoked to account for adaptation to conflict between sequential trials. To test this possibility, we capitalised on the previous observation that the absolute mean feature value of the array $|\mu|$ (i.e. proximity of the mean to the category boundary, at x = 0) impacts decisions and their latencies in a similar fashion to array variability σ^2, with no interaction between the two factors (17). In a second experiment, we again manipulated array variance (high vs. low) but varied array mean μ at 4 symmetric levels around zero (μ_2, μ_1, $-\mu_1$, $-\mu_2$ where $\mu_2 > \mu_1$). This allowed us to compare how (i) the congruity of variance (high and low) and (ii) congruity of absolute mean (μ_2 and μ_1) influenced performance. If the variance priming observed in experiment 1 is due to nonspecific effects, we would expect to additionally see priming by the absolute mean of array information. If it is specific to the range or dispersion of the information in the prime, then we would expect to replicate experiment 1 in the absence of any priming by absolute mean.

In experiment 2, we replicated the phenomenon of variance priming reported above, with faster RTs for trials on which the prime-target array variance was consistent ($F_{(1,38)} = 5.03$, $p < 0.04$). Additionally, we observed the anticipated facilitatory influence of category congruity between prime and target ($F_{(1,38)} = 20.9$, $p < 0.001$). As expected, absolute target array mean had a main effect on response times (low $|\mu|$ > high $|\mu|$, $F_{(1,38)} = 63.6$, $p < 0.001$) along with target array variance (high σ > low σ, $F_{(1,38)} = 37.8$, $p < 0.001$). Critically however, congruity in absolute mean (and thus, degree of response conflict) had no facilitatory effect on decision latencies ($F_{(1,38)} = 0.195$, $p = 0.661$) (Figure 5.2C-D, Tables A.3.1.2-3). This occurred despite this equivalence between the way that target array $|\mu|$ and σ influenced responding: the main effect of mean on RTs (low mean RT = 698 ms; high

mean RT = 679 ms difference = 19 ms) was comparable with the main effect of variance (low var RT = 678 ms; high var RT = 698 ms, difference = 20 ms). It is thus unlikely that the manipulation of mean was simply too weak to observe similar priming effects to the variance. As for experiment 1, these results cannot be accounted for by a speed accuracy trade-off, with no significant prime x target interaction on accuracy for either variance ($F_{(1,38)}$ < 0.015, p = 0.902) or absolute mean ($F_{(1,38)}$ = 1.49, p = 0.230). Together, these analyses seem to rule out nonspecific explanations based on resource recruitment or adaptation to conflict.

The influence of prime-target interval. To assess the latency with which perceptual variance influences behaviour, we collapsed across data from experiments 1 and 2, and compared variance priming at the three levels of prime-target interval (PTI) employed (100 ms, 200 ms, 500 ms). The prime-target variance interaction was significant when considering only those trials with prime-target intervals (PTI) of 100 ms ($F_{(1,78)}$ = 18.4, p < 0.001) and 200 ms ($F_{(1,78)}$ = 8.74, p < 0.004), and although the interaction failed to achieve significance for the 500 ms trials ($F_{(1,78)}$ = 2.73, p = 0.102), there was no three-way interaction between prime variance, target variance and PTI ($F_{(2,147)}$ = 1.97, p < 0.146; see Table A.3.2). This suggests that the effect begins very early and persists for several hundred milliseconds (Fig 5.3).

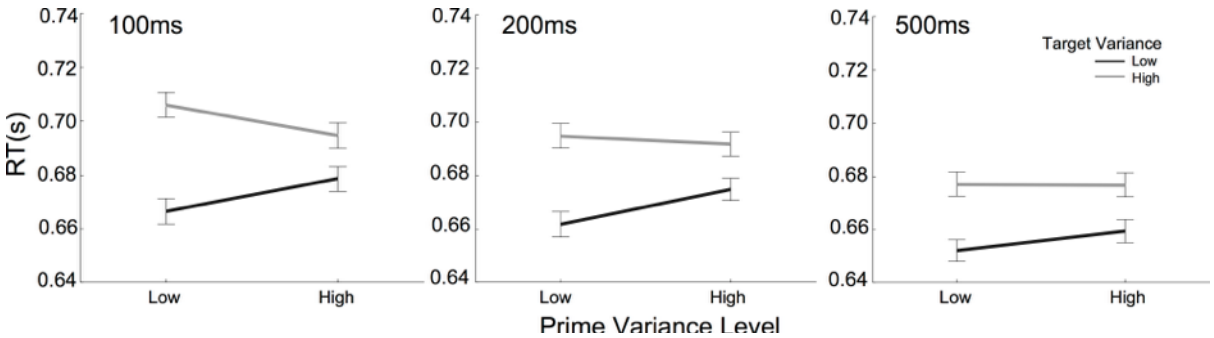


Fig 5.3. Variance priming across the three prime-target intervals (PTIs). The extent of the variance priming effect is plotted for trials grouped by the three levels of PTI, from L to R: 100, 200 and 500ms. Axis are as for Fig 5.2As.

5.2.3. Experiment 3.

If the encoding of array variability contributes to rapid detection and recognition, then one would not expect its influence to be tied to a specific spatial location. In order to test the spatial specificity of variance priming, we conducted a third experiment in which the radius of the prime and target arrays were manipulated such that either they were overlapping, or one fell concentrically within the other. A facilitatory effect of variance congruity on RT was again observed when the arrays were spatially overlapping, replicating the findings of the previous two experiments ($F_{(1,13)} = 14.2$, $p < 0.002$). The effect was again observed at both 100 ms ($F_{(1,13)} = 24.7$, $p < 0.001$) and 200 ms ($F_{(1,13)} = 4.9$, $p < 0.046$) PTI (there were no 500 ms PTI trials in experiment 3). To explore the data further, we created a summary measure of variance priming by subtracting the RT difference for each combination of congruent and incongruent different trials (high/high – high/low) – (low/high – low/low). Variance priming was significant when the prime array was large, irrespective of whether the target array was large ($t_{(1,13)} = 3.15$, $p < 0.004$) or small ($t_{(1,13)} = 2.78$, $p < 0.01$). It was also significant when the target and prime array were both small ($t_{(1,13)} = 2.48$, $p < 0.02$). Only when the prime array was small, and the target was large did the effect evaporate ($p = 0.25$) (Fig 5.4). However, neither the main effect of prime array size ($p < 0.20$), target array size ($p < 0.70$) or their interaction ($p < 0.20$) on variance priming were significant. Thus, variance priming does not depend on spatial overlap between prime and target.

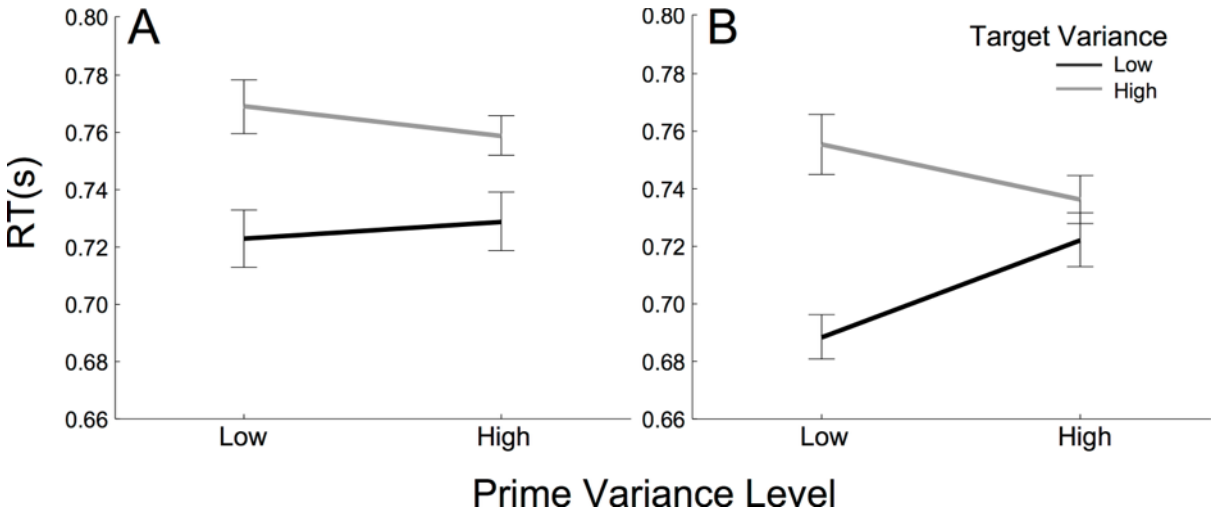


Fig 5.4. Variance priming for spatially non-overlapping prime-target pairs. Mean RTs plotted for each combination of prime variance (high vs. low) and target variance (high vs. low). Data is from Experiment 3, **(A)** for trials in which the prime array appeared within the spatial boundary of the target array and (small-large) **(B)** for the reverse spatial organisation (large –small).

5.2.4. Variance priming across the category boundary

Finally, the strongest demonstration of priming by variability would be that the congruity of prime-target variance can still facilitate responding even when prime and target are drawn from different categories (trials; e.g. red mean prime, blue mean target). We assessed this by pooling across data from all three experiments and repeating the regression analyses described above, including independent predictors for switch (prime and target from different category) and stay (same category) trials. For this analysis, we excluded data from the 500 ms condition for which variance priming was not significant (see Tables A3.4-5 for further results including all conditions). Critically, although variance priming was stronger for stay trials (prime variance x target variance x same/different category interaction, $F_{(1,92)} = 18.0$, $p < 0.001$), it was independently significant for both stay ($t_{(92)} = 7.31$, $p < 0.001$) and switch ($t_{(92)} = 1.73$, $p < 0.05$) trials (Fig 5). Individually, the effect was significant on switch trials for experiments 1 ($t_{(39)} = 1.79$, $p < 0.05$) and 3 ($t_{(13)} = 2.24$, $p < 0.03$), but not experiment 2 ($t_{(38)} = 0.69$, $p = 0.751$) (see Table A3.3 for full ANOVA results on RTs). Together, these results imply that priming by the variance of a visual array can occur entirely independently of the mean feature values of the prime and target arrays.

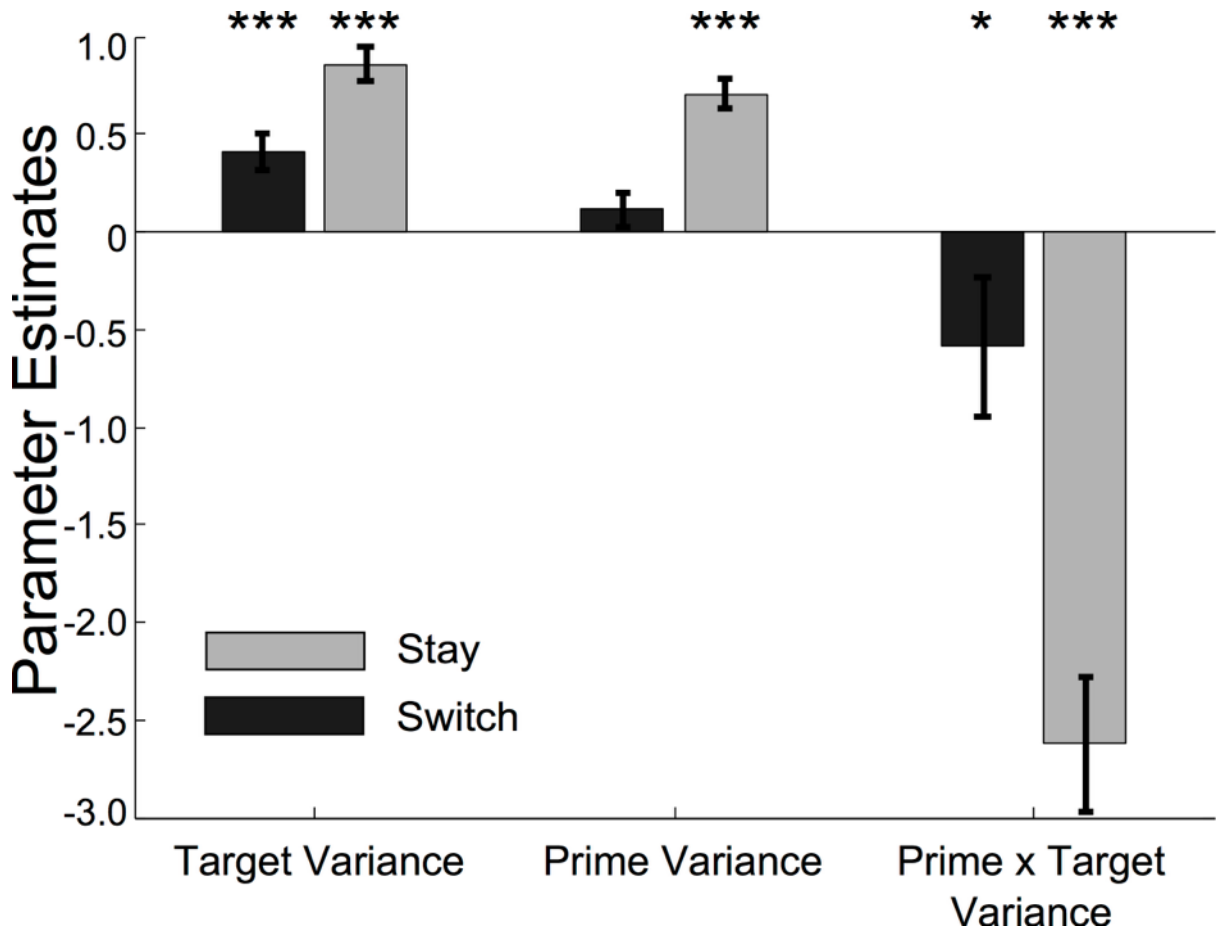


Fig 5.5. Regression weights for variance priming on category congruent (stay) and incongruent trials (switch). Regression weights are plotted for the variance of the target, prime and their interaction. Data included from the 100 and 200ms conditions across all experiments. Trials are split by whether the prime and target belong to the same category (light grey bars, 'stay') or whether the categories differ (dark grey bars, 'switch'). Stars indicate level of significance as for Fig. 2b.

Finally, we assessed the strength of the priming by variance account in the data collapsed across all experiments as described above. When all subjects were included in a single analysis, the prime variance x target variance interaction was highly significant ($F_{(1,92)}$ = 40.8, $p < 0.001$). It was also significant independently for judgments about colour ($F_{(1,52)}$ = 33.9, $p < 0.001$) and shape ($F_{(1,39)}$ = 9.21, $p < 0.004$) in roughly equal measure. Of note, the difference in RT between low/low and low/high conditions ($t_{(92)}$ = 5.96, $p < 0.001$) and high/high and high/low conditions ($t_{(92)}$ = 3.02, $p < 0.004$) were both independently significant. This latter finding rules out a ceiling effect on a response compatibility effect as an explanation for our findings.

5.2.5. Experiment 4

To further investigate the limitations of priming by variance, we conducted a fourth experiment which aimed to address whether this effect would be under the control of spatial attention. We presented participant with two simultaneous prime arrays, to the left and right visual fields. As previously, observers were then asked to categorise a single target array on the basis of its average colour. The variance of all prime and target arrays was orthogonalised. In this experiment, the target side was cued by a single white point in the centre of one of the prime arrays. If priming by variance occurs entirely automatically across all visual input then we should see priming by the combined variance of the two primes. However, if this adaptation can be controlled by spatial attention then priming should be driven most strongly by the cued prime.

Observers in this paradigm showed a robust main effect of target variance ($F_{(1,20)}$ = 17.4, $p < 0.001$) on response times. The two prime arrays were labelled as either cued (spatially congruent) or non-cued (spatially incongruent). The main effect of prime variance was not significant for either cued or non-cued prime arrays ($p > 0.10$). The cued prime variance showed a significant interaction with the target variance ($F_{(1,20)}$ = 8.42, $p < 0.01$), whilst the non-cued variance did not ($F_{(1,20)}$ = 0.051, $p = 0.483$). As previously, this interaction could not be accounted for by a speed-accuracy trade-off, with no significant interaction effect on error rates for either cued ($F_{(1,20)}$ = 2.89, $p = 0.105$) or non-cued arrays ($F_{(1,20)}$ = 0.146, $p = 0.706$), although it should be noted that there was no main effect of target

variance on error rates in this experiment ($F_{(1,20)}$ = 0.05, p = 0.828). This may be due to a ceiling effect, as performance was higher in this case than in previous experiments (mean errors for high variance = 13.8%; low variance = 14.2%). Variance priming was therefore only found for the cued prime array-target pairs.

As previously, we also used a regression approach to directly compare the influence of the cued and non-cued prime variance on the target array. We created regressors that coded the target variance, cued variance, non-cued variance and their interaction with the target variance. When entered competitively into the same analysis, the parameter estimate for the cued prime variance interaction with the target was highly significant (Fig 5.6; $t_{(20)}$ = -3.29, p<0.002), while the non-cued variance interaction did not reach significance ($t_{(20)}$ = -1.01, p=0.162). An additional possibility is that the variance of the two prime arrays is pooled, and that it is this "global" variance which drives adaptation. We therefore re-calculated the pooled variance and entered these values into a regression in competition with the cued prime variance, and their respective interactions with the target variance. Although the two interaction terms were highly correlated (r = 0.835), there remained a significant negative effect of the cued prime variance ($t_{(20)}$ = 1.92, p < 0.05) but no effect of the interaction between global variance and target variance ($t_{(20)}$ = 0.67, p = 0.25).

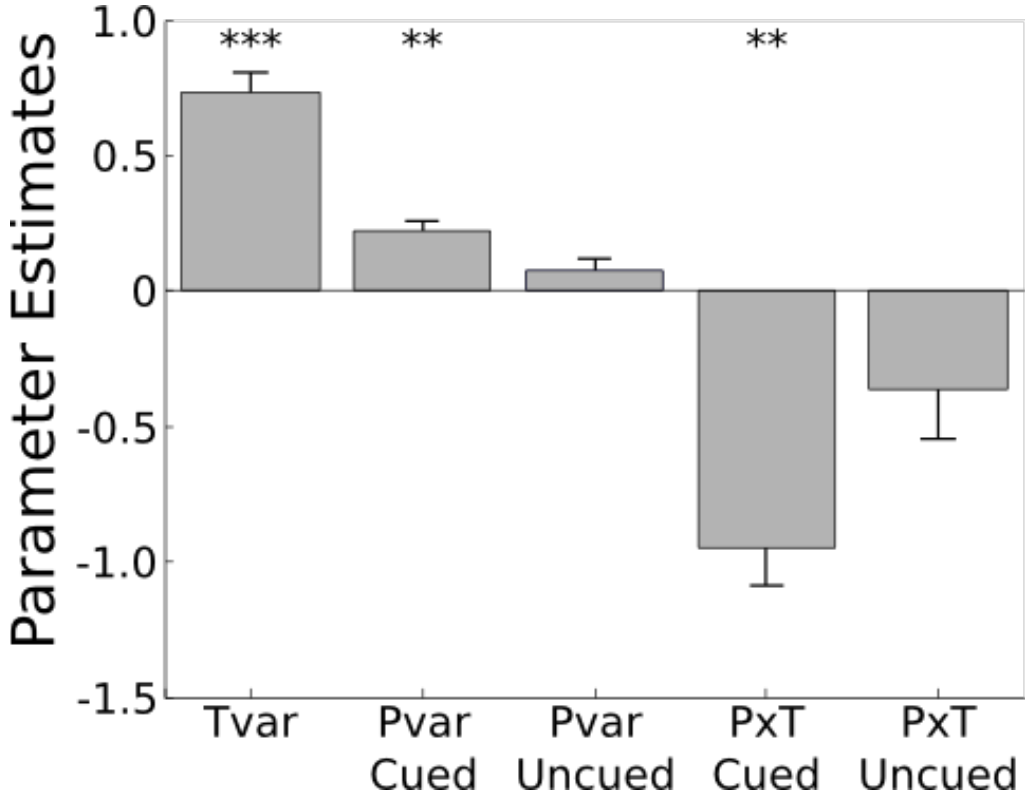


Fig 5.6. Variance Priming for Cued and Uncued Primes. Regression weights are plotted for target variance (Tvar), cued location prime variance (Pvar Cued), uncued location prime variance (Pvar Uncued), prime-target variance interaction for cued prime location (P x T Cued) and prime-target variance interaction for uncued prime location (P x T Uncued). Stars show significance level as for 2b. Error bars show SEM.

5.3 Discussion

The results presented above demonstrate a form of priming that does not rely on the repetition of specific visual features but instead is driven by the summary statistics of the display. Specifically, there is a response time facilitation when the variance of the prime and target are congruent and therefore a relative slowing in RT when the level of variance changes.

We ruled out several alternative explanations for this interaction between prime and target variance. Firstly, we could show that the interaction does not arise due to a saturation of a response compatibility effect under the high target variance condition. Secondly, there was no corresponding effect on error rates, suggesting that variance priming cannot be attributed to a form of speed-accuracy trade-off. Variance priming was also not driven by sensitivity to any increase in perceptual uncertainty, since no corresponding effect was observed for the congruence of mean distance to category boundary.

Several features of the data promote the interpretation of this effect as being driven by a summary statistic level representation. This view of representation suggests that information at the summary level is cognitively privileged and can persist in the absence of knowledge of the component features (Alvarez, 2011; Ariely, 2001). It can be contrasted with a view of perception in which individual features are the informative units of representation (Gross, Bender, & Rocha-Miranda, 1969; Tanaka, 1996). In the present case, variance priming persisted even when we took into account the difference in colour or shape between elements at the same location in space (or either side) across each prime-target pair. This shows that the effect is not due to local adaptations to fluctuations in the relevant dimension (i.e. the component features), but rather that some global summary is extracted from the whole array. However, this effect appears open to modulation by spatial attention, since knowledge of the spatial location of relevant information restricts the extraction of summary values to these locations.

Since this summary measure of variance is also category-blind, it might be expected that priming by variance should transfer across the category boundary. Indeed, although weak, we found evidence to support this in that variance priming was observed even on trials with incongruent prime-target pairs. This particularly surprising result suggests that the observed adaptive process is not just driven by the particular stimuli encountered in a given array, but that there are "system-wide" changes that affect the encoding of stimuli from the whole possible feature space (Gepshtein et al., 2013). The category congruence of prime-target pairs was also a factor that could independently drive facilitation of response times, consistent with many previous reports (Tulving & Schacter, 1990). However, since the response mappings were always consistent within a subject it is unclear whether this response facilitation is based on repeated visual category or a repeated response. Nevertheless, adaptation at the level of the categorical information appears to be an independent phenomenon from variance priming.

There is some evidence to suggest that encoding of summary variance occurs at the level of perceptual representation. Firstly, variance priming was limited to matching feature dimensions. Thus, there was a weak effect of variance congruence for the irrelevant dimension, but there was no between-dimension priming. Secondly, the rapid time course of variance priming might suggest that this occurs early in the processing hierarchy.,

This rapid time course of variance priming suggests that summary measures can be extracted very quickly from a visual input. This is interesting in the context of object and scene identification as these quantities have been suggested to underlie the ability of observers to extract high-level identify information ("gist") from complex displays, even with very limited viewing time (Biederman, 1972; Fei-Fei, Iyer, Koch, & Perona, 2007; Greene & Oliva, 2009, 2010; F. F. Li, VanRullen, Koch, & Perona, 2002). Even in the case of single objects, reports of rapid categorisation and corresponding early neural categorical signals suggest that the visual system is able to extract high level visual information is a very short space of time (Rousselet, Fabre-Thorpe, & Thorpe, 2002; Thorpe, Fize, & Marlot, 1996). The present work suggests that summary measures are available to the visual system in a time frame consistent with "gist" extraction and rapid object categorisation.

The nature of the interaction between prime and target variance is reminiscent of trial-to-trial effects in cognitive control tasks. This form of adaptation is thought to be in response to the cognitive demands of a task. For example, in tasks designed to generate varying levels of response conflict (e.g. Stroop task or Simon task), a similar facilitation of performance can be seen when the level of response conflict is matched across trials (Egner, 2007; Gratton, Coles, & Donchin, 1992). Thus, on sequential high conflict trials, performance on the second high conflict trial improves. This occurs even when no visual features are repeated between adjacent trials (Ullsperger, Bylsma, & Botvinick, 2005).

Although conflict adaptation and variance priming appear to be conceptually overlapping, it is unclear to what extent they might reflect the same phenomenon. Firstly, conflict adaptation has typically been studied in paradigms that require a response to each stimulus presentation, rather than looking at within trial adaptation (although priming paradigms have been used to investigate trial-to-trial adaptation (Kunde, 2003). Secondly, if conflict were driving the present results it might be expected that a parallel effect should have been observed for the mean distance to boundary (which can also induce increased response conflict). Therefore, it could be that variance priming and conflict adaptation reflect adaptation at different levels in the processing hierarchy.

One way to address this question would be to compare the neural representations of variance priming and conflict adaptation. Whilst conflict adaptation is thought to depend largely on dorsomedial prefrontal and lateral prefrontal regions (Kerns et al., 2004), several predictions could be made about the neural mechanism of variance priming. It could be, as neural recording in the early visual system demonstrate (Wark et al., 2007), there is adaptation at the level of the perceptual representation. Alternatively, variance priming could rely on amodal decision or control structures observed to activate in paradigms requiring sequential control of behaviour (Carter et al., 1998; Narayanan, Cavanagh, Frank, & Laubach, 2013; Sheth et al., 2012). The next chapter will aim to address these questions using an fMRI version of the present multielement priming task.

Chapter 6: Neural Correlates of Priming by the Variability of Perceptual Input

The work presented so far has been highly suggestive of a role for variance early in the visual hierarchy. Here, we aim to test this hypothesis, taking advantage of neural repetition effects to look for neural correlated of variance priming. Although both sensory and frontal regions show a clear relationship with the behavioural priming effect, this relationship is perhaps most consistent for sensory regions. Therefore, the behavioural costs of variability are likely to, in large part, rely on inefficiencies at the level of sensory representation.

As discussed in the previous chapter, efficient coding schemes predict that the gain of information processing should adapt to the statistics of the input (Attneave, 1954; Barlow, 1961). This general principle suggests that behavioural costs might arise when the resolution of the current representation does not match that required by the input. In the case of perceptual variability, a fine scale of resolution might be appropriate when the expected range of feature values is narrow and predictable, but when the variability (and thus range) increases, the same resources might need to broaden their representational range to allow for a greater variety of inputs. This broadening, however, might reduce the resources that can be devoted to making fine discriminations, leading to an increase in the potential decision time or error likelihood when such decisions need to made. The present experiment aims to investigate, using fMRI, the neural basis for these behavioural costs.

One possible source of behavioural cost might be at the level of the sensory encoding. At the neural level, single neurons are often described by their tuning preferences. However these preferences can change depending on the current demands (Fairhall et al., 2001; Scolari & Serences, 2009). Pools of neurons with more flexible coding schemes may therefore support flexible representations dependent on the recent stimulus history. Repetition of exemplars from a visual class generally leads to a reduced response rate in visual brain regions for the second and subsequent presentation compared to the first (Wiggs & Martin, 1998). For example, Desimone and colleagues showed that neurons in IT cortex show a suppression of firing following repeated exposure to images of visual objects

(Desimone, 1996; L. Li, Miller, & Desimone, 1993; Miller, Li, & Desimone, 1991). Although initially described at the single neural level (Baylis & Rolls, 1987; M. W. Brown, Wilson, & Riches, 1987), a large number of fMRI studies have reported a similar phenomenon, in that the magnitude of the BOLD signal decreases with increased likelihood of repetition (Grill-Spector, Henson, & Martin, 2006; Henson, Shallice, & Dolan, 2000; Schacter & Buckner, 1998; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008; Wiggs & Martin, 1998).

This so-called repetition-suppression is a ubiquitous property of sensory systems and occurs both when stimuli are task relevant and task-irrelevant (Miller & Desimone, 1994; Vogels, Sary, & Orban, 1995). It is a phenomenon that has attracted substantial interest due to the similarity to the behavioural observation of perceptual priming - the modulation of behavioural response dependent on the recent stimulus context. Many properties of repetition suppression are similar to the properties of perceptual priming. For example, both can occur rapidly but show a long duration, the both show graded responses relative to the frequency of repetition or exposure and additionally it is possible to observe perceptual priming and repetition suppression in response to the repetition of a particular image, even when the image is presented from different viewpoints or angles (For overview see (Grill-Spector et al., 2006; Wiggs & Martin, 1998)).

Furthermore, regardless of the mechanism of action, repetition suppression provides a useful tool for investigating which brain regions support the representation of a particular visual feature or category. For example, Kourtzi & Kanwisher (Kourtzi & Kanwisher, 2001) deployed this method to investigate the the nature of object representations in the lateral occipital complex (LOC). They constructed stimuli that could be described by their overall shape and, independently, their local contours. When these images were paired, therefore, each pair could contain repetition of contour or shape information. By comparing the extent of adaptation in these two conditions, they showed that the LOC adapted to repetitions of overall shape, but not to local contour. Therefore, this region must in some way represent the images at the level of their overall shape.

Therefore the first aim of this study is to investigate whether previously identified regions that respond to the variability of task-relevant information also demonstrate repetition-

suppression when the degree of variability is consistent between the prime and target arrays. However, given the previously described behavioural adaptation to variance repetition, we might also expect repetition suppression of the BOLD signal in response to congruent levels of variability. As described above, this form of neural modulation could indicate which neural structures are responsible for the behavioural costs associated with the level of perceptual variability. However, an alternative cognitive literature has highlighted the behavioural costs that arise from the overlap in the stimulus-response contingencies of a particular stimulus array.

This measure (termed response conflict) has been associated primarily with activity in a network of prefrontal regions; medial prefrontal cortex indexes the current level of response conflict and signals this to lateral prefrontal regions (Botvinick et al., 2001; Botvinick, Cohen, & Carter, 2004). These regions then adjust the current attentional resource distribution to better resolve any upcoming responses with a high level of conflict, resulting in improved behavioural performance for conflict-congruent trial sequences (Gratton et al., 1992). This model has a large body of supporting evidence from behavioural and neural data, but a growing body of evidence suggests that we can make conceptual and empirical distinctions between conflict arising at several levels of representation, from perception to response.

According to this view, there are additional behavioural costs that can arise when a stimulus array contains perceptually dissimilar items. For example, using modified flanker and Stroop tasks, it has been shown that there are additive effects of both stimulus and response conflict on behaviour, and that these effects may rely on different neural networks (Liston, Matalon, Hare, Davidson, & Casey, 2006; Soutschek, Taylor, Muller, & Schubert, 2013). Some have also suggested that trial to trial adaptations are primarily driven by adaptation to stimulus, but not response, conflict (Notebaert & Verguts, 2006). In the present multi element priming task, the variability of a stimulus array could be seen as a manipulation of perceptual conflict, since it increases the range of concurrent perceptual features. Therefore, we might expect that the costs of increasing variability are primarily driven by the challenge to perceptual representations.

This study will therefore use a priming version of the multi element averaging task, optimised for an fMRI environment. We expect that the behavioural adaptation observed in the previous chapter will be mirrored by a neural adaptation of the same form - reduced BOLD signal for repeated vs non-repeated levels of variability.

6.1. Methods

6.1.1. Participants

20 participants aged between 18 and 40 (7 males, 13 females) took part in this study. Subjects were compensated at a rate of 15 Euros per hour for their participation.

6.1.2. Stimuli

One each trial, participants viewed two consecutive stimulus arrays each composed of eight coloured shapes (Fig 1a). They were asked to judge the average colour of the second (target) array and ignore the immediately preceding (prime) array. The colour of each element in both prime and target arrays could vary continuously between red and blue (task-relevant feature) and the shape could vary between a circle and a square (task-irrelevant feature) To ensure continuous variation in the shape dimension, the stimuli were actually hyperelipses, a form that allows for increasing curvature between a square-like shape and a circle. For full details of the shape generation, please refer to previous chapters and (de Gardelle & Summerfield, 2011). Both the colour and the shape dimensions were parameterised on an arbitrary scale from 0 to 1, with the red/blue and circle/square boundary at 0.5. The feature values (shape and colour) on each trial were drawn from Gaussian distributions, which had a mean that could lie at a fixed interval from the category boundary in both directions. The distance of the mean from the boundary was determined by a thresholding procedure immediately prior to the scan session. The standard deviation of the Gaussian from which features were sampled was always either high (0.30) or low (0.15). Therefore, trials could be either “high” or “low” variance and the mean could fall either side of the category boundary (e.g. average red, average blue). Eight stimuli were drawn randomly from the sampling distributions for each trial, and the stimulus values were resampled to ensure they were within 0.1% tolerance of the target mean and variance. The same procedure was used to generate stimulus arrays for both the prime and target.

The eight elements in each array were displayed in a centrally presented ring around a small white fixation point. The ring was approximately 3 degrees of visual angle. Following the appearance of the central fixation point, the prime array appeared and remained on the screen for 100ms. The target array appeared 100ms after the offset of the prime (the prime-target interval, PTI). This PTI was chosen to maximise the strength of the prime as observed in previous experiments. Participants could response as soon as the target appeared, although the target only remained on the screen for 500ms (Fig 1A). Responses were made using a grip response box in the scanner.

6.1.3. Design

Each trial comprised both a prime and a target array. The absolute mean distance of the feature values from the boundary was always fixed, but this mean could fall on either side of the category boundary, meaning that an equal number of trials were drawn from each category. The key experimental manipulation was the level of stimulus variability. The variance of the prime and target array was orthogonal, and either array could be either low or high variance. Participants were only ever asked to make judgements on the colour dimension of the target array, but the same factors were also varied in the shape dimension. Therefore, the design of the experiment was 2 (prime variance, colour) x 2 (target variance, colour) x 2 (prime variance, shape) x 2 (target variance, shape).

6.1.4. Task and Procedure

Prior to the experimental session, participants underwent a thresholding procedure to determine the mean distance for both the colour and shape dimensions. Although participants were only asked to make colour judgements in the main experiment, we used this thresholding session as method for equating the behavioural saliency of the dimensions. The thresholding procedure was an adaptive staircase of 144 trials (for both colour and shape) designed to titrate performance to 80% by varying the mean feature value. This thresholding procedure was designed to be as similar as possible to the main

task, and therefore included both a prime and target array with a 100ms PTI. Response mappings were counterbalanced across participants and, for the colour task, carried through to the main experiment.

6.1.5. Behavioural Analyses

Both reaction time (RT) and accuracy data were analyses using ANOVA, using an alpha of 0.05. Analysis of RTs was performed on log-transformed, correct trials only and trials with a RT>300ms were also excluded.

6.1.6. fMRI Acquisition and Preprocessing

Images were acquired in a 3 Tesla Siemens TRIO with a 32 –channel head coil using a standard echo-planar imaging (EPI) sequence. Images were 64x64x36 volumes with voxel size 3x3x3 mm; acquired with a 2s repetition time (TR) and 30ms echo time. Two runs of 747 volumes were obtained, each of which lasted approximately 30 minutes and corresponded to one experimental block of 256 trials.

Preprocessing of the imaging data included correction for head motion and slice acquisition timing, followed by spatial normalization to the standard template brain of the Montreal Neurological Institute (MNI brain). Images were resampled to 4mm cubic voxels and spatially smoothed with a 8mm full width at half-maximum isotropic Gaussian kernel. A 128s temporal high-pass filter was applied in order to exclude low-frequency artifacts. Temporal correlations were estimated using restricted maximum likelihood estimates of variance components using a first-order autoregressive model. The resulting nonsphericity was used to form maximum likelihood estimates of the activations.

6.1.7. fMRI Analysis

All analyses were carried out using SPM8. SPM orthogonalises regressors by default, but we ensured that this feature was turned off. To separate the effects of prime and target variance on individual trials, we constructed a GLM (Model 1) with each combination of prime and target variability (i.e. low/low, low/high, high/low and high/high) modelled as independent regressors with separate onsets. For this (and all other models) only correct trials were included for analyses of effects of interest, to mirror the behavioural analysis of RTs. Error trials were modelled as a further separate regressor. Nuisance regressors were also included in this analysis to account for subject motion and scanner drift. Contrasts were generated by averaging across the relevant conditions. The main effect of target variability was thus: low/high & high/high > high/low and low/low. The main effect of prime variability was computed in a similar way: high/low & high/high > low/high and low/low. The contrast to look for correlates of the interaction term was: low/low & high/high > low/high & high/low.

A second model structure was used to account for potential cofounds or control analyses. These model used various task parameters as parametric modulators of a main effect of stimulus onset, for correct trials only. Model 2.1 used a parametric approach to account for any potential effects of the irrelevant dimension. In this case, the first regressor coded for the main effect of stimulus onset (correct trials only), with parametric modulation by: irrelevant prime std, irrelevant target std, irrelevant prime std x irrelevant target std, relevant prime std, relevant target std and relevant prime std x relevant target std. Error trials were included alongside nuisance regressors as previously. A similar approach was used to account for any residual effects of RT - in this case (Model 2.2) the 3 regressors coding for the irrelevant dimension were substituted for a regressor encoding the RT on each trial. Model 2.3 included only a single parametric modulator, which encoded the category congruency of the prime-target pair. Category congruent trials were those on which the category of the prime was identical to the category of the target.

A second line of parametric models used the a measure of similarity between prime and target pairs that was calculated from the sum of the absolute differences in colour values at each location, from the prime to the target array. This is a summary measure that

attempts to capture the extent to which there is highly dissimilar colours appearing at each element location. For this reason, we term this a measure of “element-element differences” (EED). This quantity was included as a parametric modulator of stimulus (Model 2.4) and a later analysis (Model 2.5) also included parametric modulators of the prime std, target std, prime std x target std alongside the EED regressor.

6.2 Results

All response time analyses were conducted using log-transformed RTs on correct trials only. Trials with a response time <300ms were also excluded from RT analyses.

6.2.1. Behaviour

In initial analyses, we entered the prime and target variance of the relevant dimension (colour) as factors with two levels (high vs low variance) into an ANOVA. We observed a main effect of target variance [$F_{(1,19)} = 15.2$, $p < 0.005$], with slower response times (RTs) for more highly variable arrays. More variable target arrays were also associated with increased error rates [$F_{(1,19)} = 19.1$, $p < 0.001$]. The variance of the prime did not have any significant influence on error rates or response times [RTs: $F_{(1,19)} = 0.09$, $p = 0.77$; Errors: $F_{(1,19)} = 0.73$, $p = 0.40$].There was however a significant interaction between prime and target variance on response times [$F_{(1,19)} = 13.7$, $p < 0.005$] (Fig 1b), although not for error rates [$F_{(1,19)} = 2.07$, $p = 0.17$]. To confirm that the interaction was driven equally by adaptation for low-low and high-high trials, we conducted t-tests between low-low and high-low trials [$t_{(19)} = 2.04$, $p < 0.03$] and low-high and high-high trials [$t_{(19)} = 3.49$, $p < 0.002$]. This interaction was therefore of the same form as that observed previously, with facilitation of response times for conditions with matched variance (i.e. low-low or high-high).

In addition, the variance of the irrelevant dimension was also set to be either high or low. Unlike previous experiments, however, we observed no effect of the variance of the irrelevant feature for the target [RTs: $F_{(1,19)} = 0.01$, $p = 0.938$; Errors: $F_{(1,19)} = 0.16$, $p = 0.695$], or prime [RTs: $F_{(1,19)} = 1.08$, $p = 0.311$; Errors: $F_{(1,19)} = 0.08$, $p = 0.777$] on behaviour, and no interaction [RTs: $F_{(1,19)} = 0.42$, $p = 0.521$; Errors: $F_{(1,19)} = 0.58$, $p = 0.455$].

In addition to potential variance repetitions, there was also the possibility for repetition of the stimulus categories between the prime and target arrays. The main effect of category

repetition was significant for RTs [$F_{(1,19)} = 41.1$, $p < 0.001$], such that responses to the target were facilitated by congruent prime-target pairs (i.e red-red or blue-blue), but there was no such effect for error rates [$F_{(1,19)} = 0.02$, $p = 0.893$]. We had also previously reported that variance priming on the relevant dimension was significant even when there was a switch in category between prime and target arrays. In the current data set, variance priming was highly significant for category congruent trials [$F_{(1,19)} = 14.6$, $p < 0.002$] but did not reach significance for category incongruent trials [$F_{(1,19)} = 0.49$, $p = 0.493$].

We also again tested whether the observed priming by variability could be explained by the local differences, instead of at the summary level. We therefore calculated the absolute difference between the feature values at each array location for each prime-target pair (element-element difference, EED). The sum of these values provided an alternative measure of the similarity of the prime and target, but is calculated at the level of local feature differences. To assess whether EED could account for the behavioural priming effect, we entered EED as a predictor of RTs in a regression analysis, alongside the prime variance, target variance and the prime-target interaction term. One sample t-tests of the group parameter estimates showed that the EED have a significant main effect [$t_{(19)} = 1.95$, $p < 0.05$], such that RTs are slowed for increasing differences. Importantly, however, the parameter estimates for the interaction term remained highly significant [$t_{(19)} = 3.22$, $p < 0.01$]. Therefore, the priming effect is driven by the summary-level variability, rather than a measure of local feature differences.

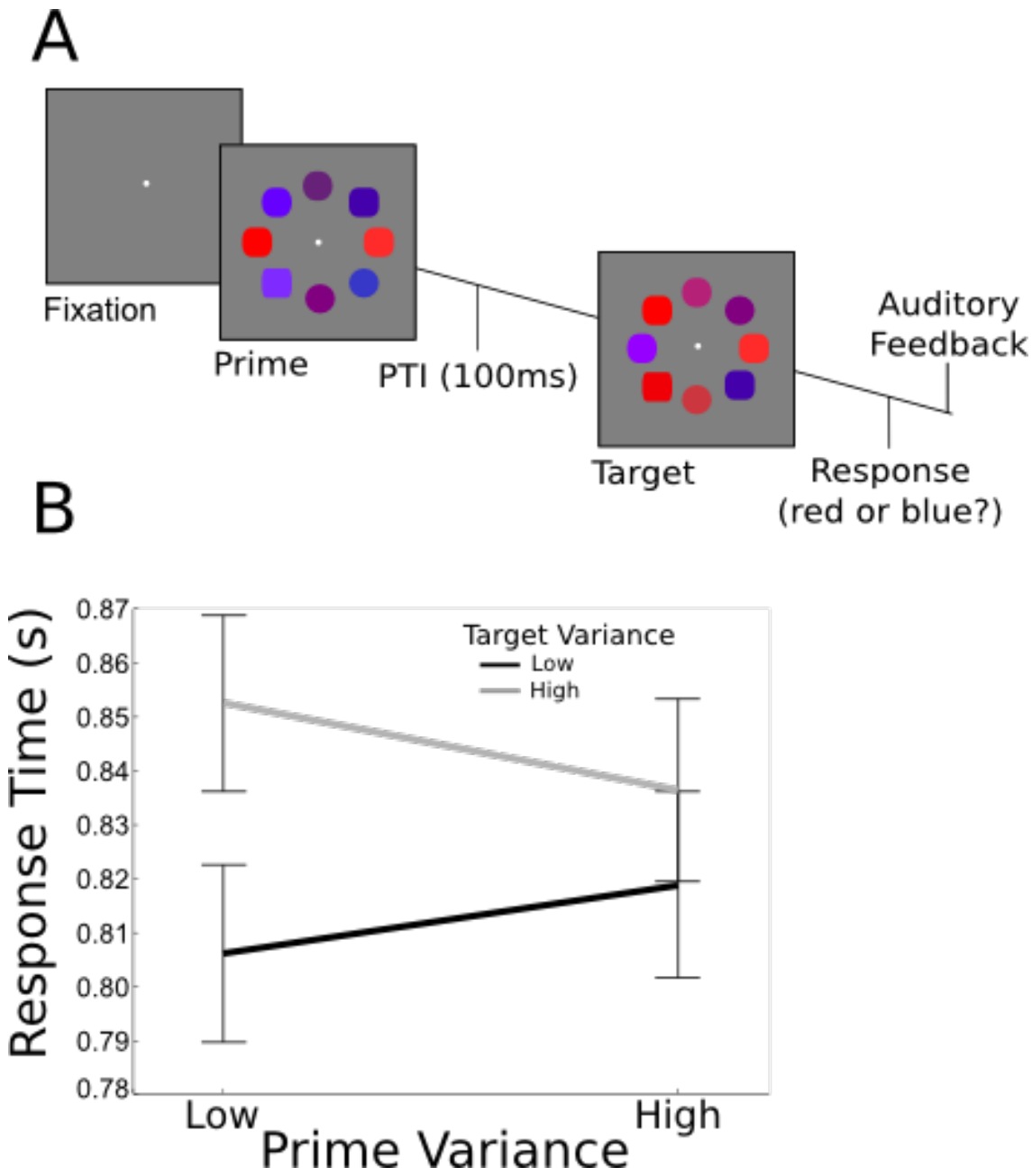


Fig 6.1. Task and Behaviour. A. Behavioural Task. Human observers were asked to categorise a target array of eight multicoloured shapes into either “red” or “blue” categories, dependent on the average colour of all elements. A task-irrelevant prime array preceded this target, which also contained eight multicoloured shapes with an average colour that was either blue or red. The prime was displayed for 100ms and there was a 100ms interval (Prime-Target Interval; PTI). between the prime and target arrays. Auditory feedback followed response. B. Effect of prime and target variability on response times (RTs). X-axis plots the effect of prime variability (low or high), whilst the lines show effect

of target variability which was either high (light grey line) or low (black line). Error bars show SEM.

6.2.2. Neural Effects of Target Variance

Before addressing any modulatory effect of prime variance on BOLD signals, we first looked at whether the main effect of target variance followed the previously observed pattern (Chapter 5). We entered each prime-target variance combination (i.e. low/low, low/high, high/low, high/high) as independent regressors into a GLM (Model 1, see Methods). Only correct trials were included in this analysis, error trials were modelled separately. Unless stated otherwise, all reported statistics are significant at $p<0.05$ with cluster correction applied (FDR).

The main effect of target variance broadly replicated the previously observed pattern of activation (Fig 6.2). The global peak was centred in left visual cortex [left: -34, -80, 6; $t(19) = 6.94$, $p<0.001$, Fig 2a], where high variance was associated with a greater BOLD response. A similar positive correlation was observed bilaterally in a premotor region [right: 26, -8, 50, $t(19)=4.88$, $p<0.001$; left: -26, -12, 54 $t(19) = 4.40$, $p<0.001$ Fig 2b] and in right lateral prefrontal cortex [46, 32, 18, $t(19) = 4.38$, $p<0.001$ Fig 6.2c]. Unlike the previous experiment, however, we found no negative activation in medial prefrontal regions. Instead, we observed a modest positive correlation between variability and BOLD response in this region [14, 16,40 $t_{(19)} = 4.30$, $p<0.001$], although this region did not survive cluster correction. Tables of suprathreshold activations can be found in Appendix 4.

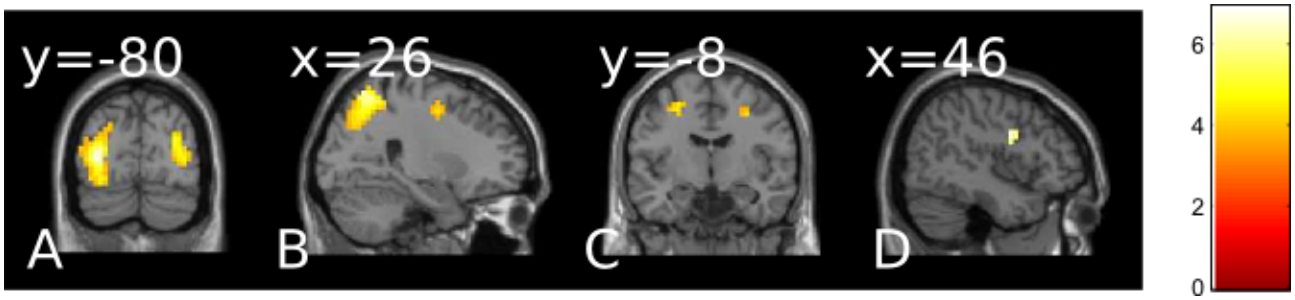


Fig 6.2. Main Effect of Target Variance. All images show positive correlations between target variability and BOLD signal in A) Left visual cortex B) Left parietal cortex C) Bilateral premotor cortex D) Right lateral prefrontal (lPFC) Cortex. Images thresholded at $p<0.001$, uncorrected but all regions retain significance at $p<0.05$ following correction for multiple comparisons. Colour bar indicates range of t-values.

6.2.3. Priming by Variability in the BOLD signal.

The main question for the present study was to identify potential neural correlates of the adaptation to variance observed in behavioural data. To identify such regions, we created a GLM which contained regressors for prime variance, target variance and their interaction term as parametric modulators of the main effect of stimulation. As described above, increasing target variance induced a greater BOLD response across a wide range of posterior brain regions. In contrast, no voxels showed a significant association between BOLD signal and the level of prime variability. The interaction term, however, was associated with several regions of negative activation (Fig 6.3, Table A4.1.2). This confirms our hypothesis that the interaction between prime and target variability would manifest as a repetition suppression effect: the BOLD signal was greatest when the variability was inconsistent across prime-target pairs. These regions were centred in left ventral temporal cortex [-54, -48 -10, $t_{(19)} = 5.35$, $p < 0.001$, Fig 6.3a], right premotor cortex [26, -8, 54 $t(19) = 5.00$, $p < 0.001$, Fig 6.3b] and right lateral prefrontal cortex [34, 24, 18 $t(19) = 4.07$, $p < 0.001$, Fig 6.3c]. No region showed a positive relationship between the interaction term and BOLD signal.

To confirm the nature of the interaction driving the neural responses in ventral temporal, lateral prefrontal and premotor regions, we extracted the parameter estimates associated with each combination of prime and target variance (plotted underneath each ROI in Fig 6.3a-c). Barplots of the extracted t-values across subjects show that the interaction in these regions takes the same form as for RTs; the most negative activation is observed when the level of variance is repeated from prime and target. For all three ROIs, there was a significant difference between low-low/high-low conditions [Temporal: $t_{(19)} = 2.58$, $p < 0.05$; Premotor: $t_{(19)} = 2.91$, $p < 0.05$; dlPFC: $t_{(19)} = 4.65$, $p < 0.001$]. However, the low-high/high-high difference was only significant for the temporal and premotor regions [Temporal: $t_{(19)} = 3.39$, $p < 0.01$; Premotor: $t_{(19)} = 2.61$, $p < 0.05$].

It might be expected that regions responding to the interaction term would also show sensitivity to the prime or target variability. Figure 6.4 plots the mean parameter

estimates for the main effects of prime and target variability in a region around the peak voxel in the three ROIs showing negative responses to the interaction term. These negative correlations are plotted for comparison. Although all regions show a positive main effect of target variability, this only reaches significance for the lateral prefrontal [$t_{(19)} = 1.80$, $p < 0.05$; Fig 6.4c] and premotor ROI [$t_{(19)} = 4.05$, $p < 0.001$; Fig 6.4b]. No region showed a significant main effect of prime variability. Together, these results suggest that regions sensitive to the prime-target variance congruency are partially-overlapping with those showing a main effect of target variability.

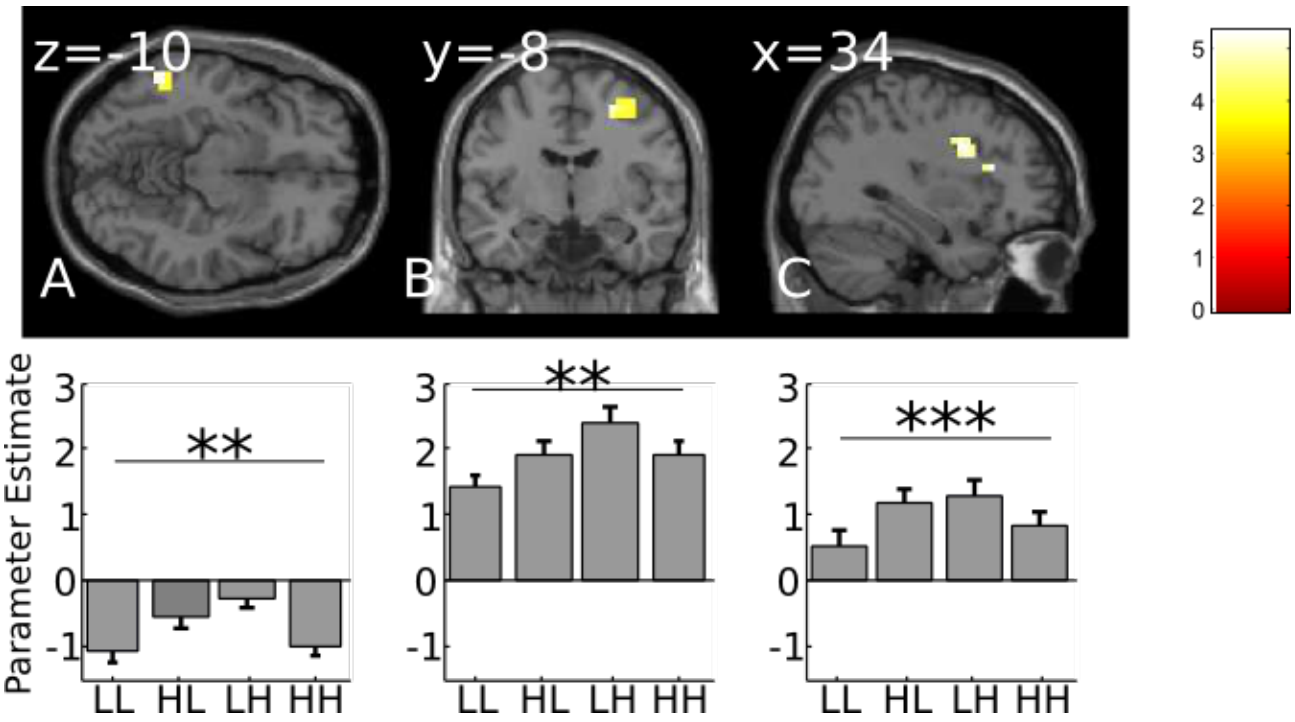


Fig 6.3. Neural Correlates of Variance Priming. All neural plots show reflect a negative correlation between the BOLD signal and the magnitude of the interaction term in A) Ventral temporal cortex B) Premotor cortex C) Lateral prefrontal cortex. Colour bar indicates range of t-values. Bar plots show the parameter estimates for each possible prime-target variance combination (x axis - LL = low/low, HL = high/low, LH = low/high, HH - high/high). Starts indicate the significance level of a repeated measures ANOVA across the four conditions: ** = $p < 0.01$, *** = $p < 0.001$.

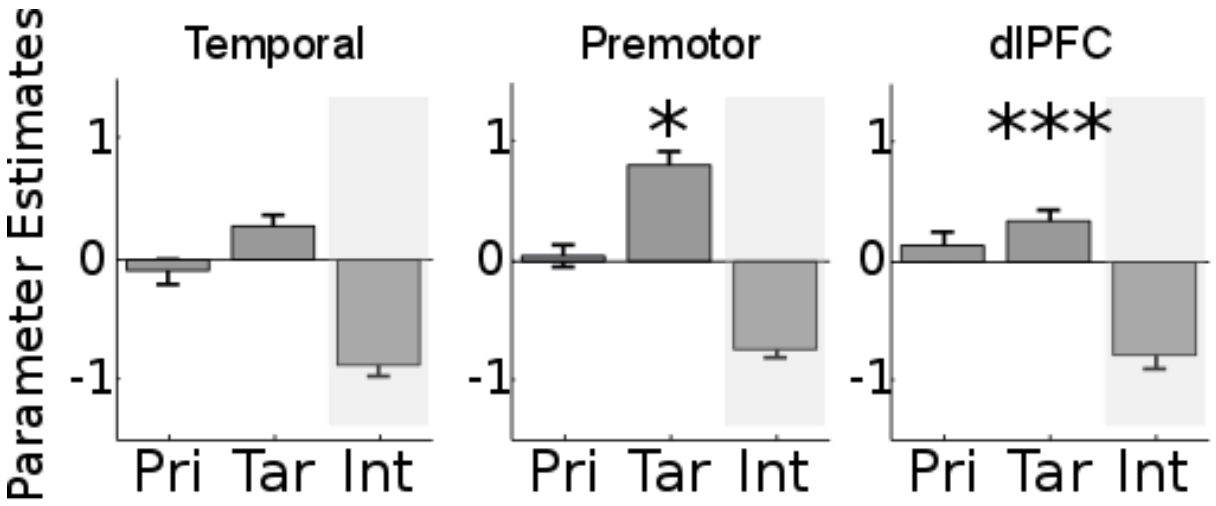


Fig 6.4. Bar plots of Main Effect and Interactions in Priming ROIs. Barplots of parameter estimates associated with the variability of the prime array (Pri), target array (Tar) or their interaction (Int). Estimates were extracted from ROIs selected on the basis of their significant negative correlation with the interaction term, so these values (with greyed out bars) are presented for comparison purposes only. Error bars show SEM. Stars indicate significance of a one-sample t-test: * = $p<0.05$, *** = $p<0.001$

6.2.4. Irrelevant vs Relevant Dimension

To determine whether the activity related to the prime and target variance was decision-relevant or instead reflected some general processing demand, we conducted a similar analysis as described above but included the variability of the prime and target on the irrelevant dimension, as well as the interaction term for the irrelevant dimension (Model 2.1). Some areas of visual cortex showed a positive correlation with the main effect of irrelevant variability [left: $t_{(19)}$ = 4.90, -26 -76 2; right: $t_{(19)}$ = 4.67, 30 -68 14]. Right lateral prefrontal cortex also responded positively to increasing irrelevant variability [38, 28 18, $t_{(19)}$ = 4.38]. No significant clusters were observed to correlated with the prime or interaction terms for the irrelevant dimension, suggesting the previously observed effects are specific to task or decision relevant information.

Since these terms were entered as competitive regressors for the task-relevant effect, we can assess whether the previous effects remained significant once the effects of the irrelevant dimension were taken into account. This is a more direct test of the independence of effects of the two dimensions. The main effect of target variability was, as previously, positively associated with BOLD signal in right premotor [$t_{(19)}$ = 5.83, 26 -8 54], and right parietal cortex [$t_{(19)}$ = 5.13, 26 -56 50]. Bilateral ventral temporal cortex also showed a positive main effect of variability [left: $t_{(19)}$ = 5.81, -54 -52 -10; right: $t_{(19)}$ = 5.19, 42 -60 -6]. The variability of the prime array was also positively correlated with the BOLD signal in a right lateral prefrontal region [$t_{(19)}$ = 5.75, 34 24 18]. The previously observed negative correlations with the interaction term were still present in right premotor cortex [$t_{(19)}$ = 5.18, 26 -8 54] and ventral temporal cortex [$t_{(19)}$ = 5.39, 54, 48 -10], although they showed a weakened effect that did not quite reach significance at the cluster-corrected level [both $p < 0.08$]. However, using an ROI approach (regions defined as shown in Fig 6.3), the parameter estimates for the relevant dimension interaction term remained significant in this model [temporal: $t_{(19)}$ = 4.45, $p < 0.001$, premotor: $t_{(19)}$ = 4.63, $p < 0.001$, dlPFC: $t_{(19)}$ = 3.98, $p < 0.001$]. Therefore, the neural correlates of variance priming are specific to the decision-relevant information.

6.2.5. Inclusion of RT.

Since the behavioural effects are manifested most clearly in response times, we also included response times (RTs) as a regressor alongside prime and target variance and the interaction term (Model 2.2). This also acts as a form of control analysis to determine whether regions responding to the interaction term are just responding to the modulations of response times, or whether these regions hold adaptive representations that subsequently drive changes in response times. RTs were strongly positively correlated with BOLD activity in medial prefrontal cortex [$t_{(19)}$ = 8.12, -6 16 50] and bilaterally in parietal cortex [left: $t_{(19)}$ = 4.85, -22 -68 50; right t(19) = 6.14, 38 -44 46], commensurate with previous reports and with the positive relationship these regions show with increasing variability. Although the pattern of activation was preserved for the interaction term, the ventral temporal [-54 -48 -10, $t_{(19)}$ = 5.24] and lateral prefrontal [34 24 18, $t_{(19)}$ = 4.83] regions did not reach significance at the cluster corrected level [ventral temporal: p = 0.16; lateral pfc: p < 0.29].

6.2.6. Category Congruency Effects.

In addition to facilitation of RTs for conditions of matching variance, RTs were also facilitated by the repetition of the same category for prime and target arrays. Although this category priming is independent from variance priming, it is possible that they both arise from changes in the same underlying neural representations. To test this possibility, we constructed a GLM (Model 2.3) with a parametric regressor which classified each trial as either “switch” (e.g. blue prime- red target) or “stay” (blue prime-blue target). We would expect that a brain region involved in categorical priming would show greater activation for switch trials relative to stay trials. However, no significant neural differences in any region were found for switch relative to stay trials.

6.2.7. Accounting for Element-Element Differences in the Neural Data

An alternative measure of categorical congruency is the sum of the differences between individual elements from prime and target arrays that fall in the same spatial location. When the two arrays are drawn from the same category, individual elements from the prime and target arrays that fall at the same spatial location are more likely to be similar. However, this will vary depending on the local differences. Therefore, the sum of the absolute local differences (element-element differences, EED) can provide a parametric measure of categorical similarity between the prime and target arrays. We therefore entered this measure as a parametric modulator of the main effect of stimulation (Model 2.4). The results show that a wide range of regions correlate with this measure (Fig 6.5a-c), including parietal [left: -26 -60 58, t(19) = 5.34; right: 34 -56 58, $t_{(19)}$ = 6.82], premotor [left: -22 -4 54, $t_{(19)}$ = 6.30; right: 30 0 54, $t_{(19)}$ = 8.46] and right lateral pfc [54 8 22, $t_{(19)}$ = 5.30] regions.

The EED also provides a measure of local variability, a quantity that could potentially be driving the observed adaptations. We therefore entered these differences into a GLM alongside the prime variability, target variability and the prime-target interaction term (Model 2.5). This analysis allows us to determine whether the neural adaptation is truly driven by a summary level quantity, or whether it can be accounted for by local feature repetition. We extracted parameter estimates from this model for a temporal region sensitive to the main effect of EED at a reduced threshold of $p<0.005$. Fig 6.5-D shows that, in the temporal ROI, there was a significant negative effect of the interaction term [$t_{(19)}$ = 2.75, $p<0.005$]. For comparison, in the premotor and dlPFC EED ROIs defined above, the interaction term was significant for premotor cortex [$t_{(19)}$ = 2.53, $p < 0.02$ but not for dlPFC [$t_{(19)}$ = 1.24, $p = 0.885$]. Therefore, the neural effect of variance congruence cannot be accounted for by differences at the local level in any of the three ROIs.

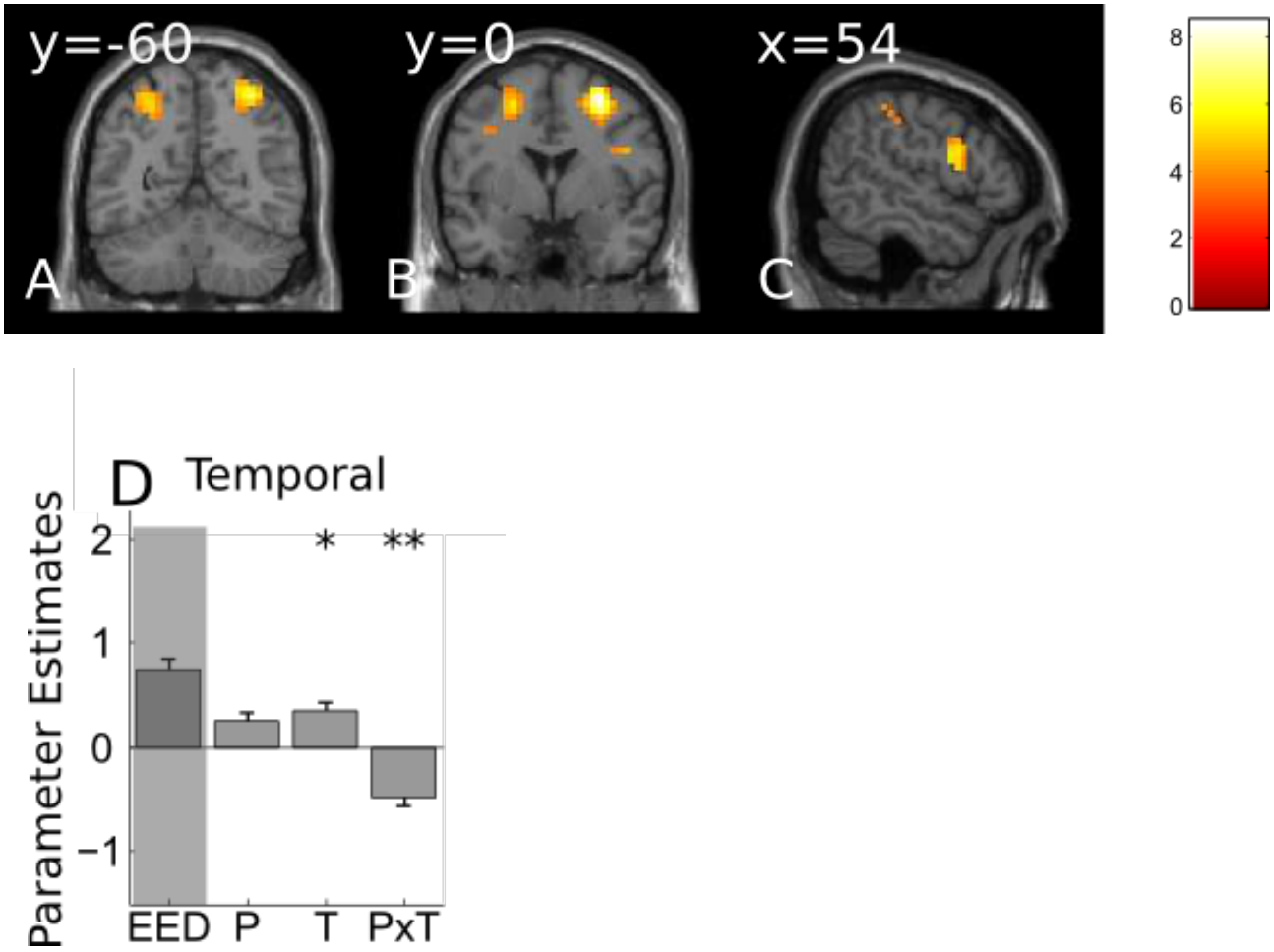


Fig 6.5. Main Effect of Element-Element Differences (EEDs). Plots A-C all show positive correlations between BOLD signal and increasing element-element differences in A) bilateral parietal regions B) bilateral premotor cortex and C) right lateral prefrontal cortex. Colour bar indicates range of t-values. Images thresholded at $p<0.001$, uncorrected all results survive cluster correction at $p<0.05$.

Barplots D show parameter estimates extracted from a temporal ROI responsive to element-element differences at $p<0.005$ uncorrected. X-axis labels: element-element differences (EEDs), prime variance (P), target variance (T), and the prime variance x target variance interaction (P x T). Stars indicate significance of a one sample t-test, * = $p<0.05$, ** = $p < 0.01$

6.3 Discussion

The data presented above explore the neural correlates of the behavioural facilitation observed when the variance on a current target array is matched by a previously seen array. We replicated the previous behavioural observation that performance is facilitated when the variance of the prime and target arrays is matched, irrespective of whether the overall variability is high or low. We extended these findings here to show that this effect is reflected in the predicted repetition suppression for variance congruent conditions. These results suggest that several brain regions contribute to the behavioural cost associated with increasing variability.

The main effect of target variability was, as seen previously, reflected in a positive correlation between increasing variability and BOLD signal. This relationship was observed across several brain regions, including parietal, visual and lateral prefrontal cortices. The increase in BOLD activation with increasing variability is consistent with the increased processing time associated with variable arrays. Since this pattern of activation was not observed for the prime array, it seems to reflect decision-relevant activity rather than an automatic sensory response. Furthermore, large portions of this activation pattern remained significant even once the variability of the irrelevant dimension was taken into account. This suggests that these regions were not just responding to any source of perceptual variability, but that this activation reflected the processing of a selected subset of information.

However, the primary aim of the current project was to look for brain regions modulated by the congruency of the prime-target array variability. Although the variability of the prime showed no significant neural effect, prime target pairs with consistent variability were associated most consistently with reduced activation across a ventral temporal and a lateral prefrontal region. This direction of the effect was exactly as predicted and is a form of repetition suppression, since there was reduced activation for a repeated stimulus feature. Unlike many studies of repetition suppression in the visual domain, however, the repeated feature in the present case is a summary value that described the distribution of information across the relevant feature space. This provides further evidence that

summary information is encoded by sensory systems and is influential in perceptual categorisation when multiple sources of information need to be integrated.

It is important to note that the present results do not argue for “variance selectivity”. Rather, we would suggest that they support a scheme in which increased perceptual variability results in increased activation of the sensory structures that are necessary for task performance. In addition, these neural populations adapt to best match their sensitivity to the current range of sensory input, potentially via a mechanism such as gain control. Therefore, one contribution to the behavioural costs associated with variability might be the mis-match between the current representational state and the likelihood of inputs at different points along the possible sensory range.

A second question this study aimed to address was whether the previously observed behavioural priming was associated with prefrontal regions often associated with cognitive control(Botvinick et al., 2001; Botvinick et al., 2004). The current results do demonstrate the involvement of lateral prefrontal areas in the adaptation to congruent variability. However, this region did not demonstrate a significant association with the strength of the behavioural effect. This might suggest that lateral prefrontal regions are more broadly sensitive to perceptual variability, whilst the ventral temporal region is, by virtue of the selectivity of visual neurons, limited in its response to the task-relevant information.

These results support a distinction that has been drawn between conflict occurring at the response level or at a perceptual (stimulus-based) level. Evidence from fMRI suggests that mPFC regions are limited to a role in response conflict (Milham & Banich, 2005; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001), whilst more posterior regions may support the resolution of perceptual conflict ((Liston et al., 2006)). Causal evidence from TMS also supports this distinction (Soutschek et al., 2013). In this latter study, TMS evidence suggests that the resolution of perceptual and response level conflict occurred in series. Thus, any effect of perceptual conflict was added to that of response conflict. Therefore, in the present results the activity in more prefrontal and premotor regions could represent the downstream effects of perceptual conflict resolution, but in a motor or response representational frame.

The temporal ROI is likely to therefore represent a perceptual representation that is of direct importance for task performance. However, it is unclear exactly which quantity or visual feature this region is sensitive to. One possibility is that it reflects high-level visual identity information, such as observed in the nearby LOC (Kourtzi & Kanwisher, 2000, 2001). Furthermore, the few fMRI studies to investigate the processing of ensembles of objects have demonstrated the involvement of ventro-medial temporal regions (Cant & Xu, 2012). This study showed that this region adapted across repetitions of object ensembles, in contrast to LOC, which only showed adaptation to repeats of the same objects. Therefore, there appears to be a hierarchy in the ventral visual stream that moves from single object identity to representation of configurations, perhaps for the discrimination visual texture in the environment (Beason-Held et al., 1998). Although the temporal ROI in the present data does not overlap with the "texture region", these results might suggest that sensitivity to the higher order statistics of the input may be a useful function for a region in this position in the object identification hierarchy.

One discrepancy between the present behavioural and neural findings concerned the effect of category repetition. As described in the introduction, the repetition of exemplars from the same category will usually lead to a repetition suppression for the repeated item (Tulving & Schacter, 1990; Wiggs & Martin, 1998). However, no such effects were observed in the current data, despite a very strong facilitation of RTs for repeated vs non-repeated categories for prime-target pairs. One possible explanation for this is that the adaptation to category relies on effects that occur at the level of individual elements . Although several studies argue that calculation of summary statistics can occur without the storage of any constituent featural information (Ariely, 2001), very few of these studies look at adaptive effects. Interestingly, our measure of local category congruency ("element-element differences") was quite highly correlated with the variability. This might suggest that the lack of neural effects of category repetition may be masked by the stimulus variability.

However, the fMRI results presented here provide a neural correlate for the behavioural adaptation to variability described in Chapter 5. Several brain regions show an interaction

between prime and target variability of the same form as the behavioural data, including a ventral temporal region and regions of sensory cortex with strong association with the representation of task-relevant stimulus features. Furthermore, this adaptation reflects processing that occurs at the summary level, providing further evidence for encoding of summary representations by the visual system.

Chapter 7: General Discussion

This final chapter will draw together the results across the four sets of experiments and discuss the overarching themes and conclusions.

7.1. Dissociable sources of perceptual uncertainty.

The work presented here aimed to investigate the behavioural and neural dissociations between two sources of perceptual uncertainty. When categorical choices are made on the basis of multiple sources of information, this information needs to be combined before a choice is made. Although, in theory, the only relevant quantity for observers is the mean feature value, the present results demonstrate that the variability of the information under consideration has a strong influence on behavioural and neural indices of the decision process. Independent influences on decisions made under uncertainty is in itself not a new idea. In the literature on value-based decision making, quantities such as risk and reward independently contribute to choice (O'Neill & Schultz, 2014) and seem to have independent neural representations (O'Neill & Schultz, 2010). In the perceptual domain, theoretical models have long demonstrated that the variability and separability (in this case, distance from the category boundary) of information should independently influence observer sensitivity (Green & Swets, 1966). However, this distinction has rarely been drawn in empirical studies of perceptual choice.

Both increasing variability and decreasing mean distance of the feature values to the category boundary are associated with increased error rates and prolonged response times (RTs), replicating a previous set of findings (de Gardelle & Summerfield, 2011). One important finding from the previous work was that behaviour could be best described by a model (LPR model) in which the evidential value of individual elements was transformed by a sigmoid function. The shape of this function is not assumed a priori; it results from the mixture of the Gaussian distributions of feature values. The shape of the transfer function therefore emerges from the structure of trials in this task. The resulting “decision space” values reflects the likelihood of a response given the feature values of the array

under consideration. The primary consequence of this transformation is that values that lie further from the centre of the trial distribution (extreme values) are underweighted compared to those that lie towards the centre (inlying elements). The behavioural results presented here were also shown to be best described by such a model, and subjects' behaviour indicated that they also showed the characteristic down weighting of outlying elements.

Further confirmation of this model was provided in Chapter 3 and Chapter 4, which tested whether the LPR model could account for neural activity as measured with fMRI and EEG. In both cases, the LPR model was able to account well for the neural data. Chapter 3 demonstrated, in fact, that the LPR model produced more robust correlations between the two measures of uncertainty and BOLD signal. These data also demonstrated that the brain regions modulated by the array statistics (i.e. mean or variance) were partially independent. There were areas of convergence in the parietal cortex, consistently implicated in the representation of a decision variable (Fitzgerald et al., 2012; Gold & Shadlen, 2007; Heekeren et al., 2008). It might therefore be expected that this region should be sensitive to all decision-relevant sources of uncertainty, since they influence the development of the decision variable. Large portions of visual cortex were sensitive to increasing variability, but seemed insensitive to changes in the mean distance to category boundary (for either decision space or native space values). This was replicated in a second fMRI experiment (Chapter 6), and was perhaps also reflected in the modulation of an early ERP (Chapter 4). Note that it is difficult to draw strong conclusions about whether this activity reflects variability in native space or decision space values, since the correlation between the two is high. Modulations by the decision space mean (U_m) were also observed uniquely in lateral prefrontal cortex, although a similar region was later found to be sensitive to variability (Chapter 6), suggesting that perhaps this region is modulated by multiple sources of decision-relevant uncertainty, but in a manner that depends on the task parameters.

The most interesting representation of uncertainty was observed in medial prefrontal cortex. In this case, there was a positive correlation between BOLD and U_m, but a negative correlation with U_v. This is clearly in opposition to what would be predicted for this region

based on many studies highlighting the role of this region in conflict monitoring (Botvinick et al., 2001; Carter et al., 1998; Yeung et al., 2004) and reflections of error likelihood (J. W. Brown & Braver, 2005). These theories would predict a positive relationship for both U_m and U_v, since both are associated with increasing error likelihood and response conflict. It could be that our results reflect the activity of two populations of neurons; some sensitive to U_m and some sensitive to U_v. This question could only really be addressed by experiments at the single unit level, but the finding that the positive U_m/negative U_v relationship was observed within individual voxels makes this coding structure more unlikely. In search of a more parsimonious explanation, we instead suggest that the mPFC relationship with uncertainty could reflect the expression of the LPR model. This model states that the greatest discrimination sensitivity should occur at the region closest to the category boundary. Changes in neuronal sensitivity could be implemented by a change in the gain associated with neurons representing this region of feature space. Gain control is a useful mechanism for rapid and flexible modulation of the influence of neuronal outputs (Reynolds & Heeger, 2009; Scolari & Serences, 2009), without recourse to long term plastic changes. For the LPR model, the input that would be associated with the highest gain would be a series of elements clustered tightly around the category boundary. In the current task parameterisation, this would be equivalent to a low mean, low variance trial - exactly those trials associated with the highest level of activation in mPFC.

In Chapter 4, we replicated the opposing modulations of the neural signal by U_m and U_v. The direction of the effects was the same; the EEG amplitude was greatest on those trials with a low mean and low variability. This effect was most pronounced at central midline electrodes, which would be consistent with a mPFC source. Interestingly, this relationship was observed during the decision phase of the trial, consistent with an account in which the level of gain applied to the output of neurons representing the elements in feature space is reflected in neural activity prior to response. This account is also not necessarily at odds with several other prominent theories of mPFC function. For example, a large body of evidence suggests that the mPFC learns the value of actions (Behrens, Woolrich, Walton, & Rushworth, 2007; Rudebeck et al., 2008) to guide value-base choice. In the perceptual

domain, this could translate to learning which portions of feature space are the most informative to determine the identity of a visual input.

However, although the evidence presented in Chapter 3 and 4 are consistent with this explanation, they were not designed to explicitly test this gain control hypothesis. Further work would be needed, for example, by monitoring the mPFC response to uncertainty (i.e U_m and U_v) as the category boundary changed position in feature space. In this case, we would predict that mPFC activity would track not the absolute variability or mean distance to boundary, but the overall distance of the elements from the observer's estimate of boundary location. Furthermore, it is not clear whether the mPFC signal reflects the direct expression of changes in gain or whether this is a form of control signal which acts to modulate the output of an earlier brain region.

Another issue yet to be resolved is that the negative activation of mPFC was not replicated in Chapter 6. This is perhaps due to the different nature of the task; perhaps the control applied by mPFC is expressed differently when there is an obvious distractor that needs to be ignored or there are rapid adaptations such as variance priming. Additionally, it could be that the gain control mechanism is not expressed as strongly under this task parameterisation, since this task employed a fixed mean distance from boundary. This would perhaps be a question for future studies.

7.2. Relationship to Summary Representation.

One literature to which this work contributes is the growing interest in summary representations. This form of representation is argued to be qualitatively different to representation by a collection of individual features (Alvarez, 2011). Behaviourally, human observers seem to retain knowledge of the average feature of a set of objects without (accessible) knowledge of the component parts (Ariely, 2001; Marchant & de Fockert, 2009), suggesting to some that the summary level reflects a cognitively privileged level of representation.

The present work also demonstrates the ease with which observers can compute a summary value of a set of heterogeneous items. Although there has been some concern over the use of sub-sampling strategies, the weighting of elements by their location in feature space (as predicted by the LPR model) suggests that, at the very least, individual elements are represented at the basic featural level first. It is possible that a subset of these elements are then selected for further integration into the final decision process (Myczek & Simons, 2008), although evidence from other tasks suggests that a sub-sampling strategy cannot account for performance in many averaging tasks (Ariely, 2008; Chong et al., 2008).

It should also be noted that the representation of mean distance to boundary in the current experiment is the least surprising, since this is the quantity that enabled observers to perform the task. It is also not clear whether the summary value is represented prior to, or as a result of, the categorisation process. Models of sequential sampling might assume in this case that each sample would represent the LPR value associated with each individual element. However, if summary values are extracted at a rapid, pre-attentive stage of visual processing, then the decision process could repeatedly sample from this internal summary representation. Perhaps one way to dissociate automatic summary representation from response-relevant processing would be to ask observers to judge the variability, rather than the mean, of visual arrays. In this case, the mean would become irrelevant and any residual effects of mean distance to boundary would suggest that there is some automatic summary representation.

In the current experiments, however, the variability of the array was always task-irrelevant but nevertheless influenced decision accuracy, RTs and was associated with an increase in BOLD signal across large portions of visual cortex. Perhaps the strongest evidence for automatic summary representation is presented in Chapters 5 and 6. In this case, the variability of a task-irrelevant prime array was shown to modulate RTs associated with the categorisation of the target array. A strong interpretation of these data might suggest that representation of these summary values implies representation of the full distribution of perceptual input. However, it could be that these summary values are estimated by a more heuristic approach. For example, the variance could be approximated by calculation of the

range. Since the experimental designs to not sufficiently dissociate variability from other measures, such as the range, it is not possible to conclude which quantity is driving the neural and behavioural results.

7.3. How might mean and variability be represented by the neural code?

Although summary values are clearly represented by several brain regions, the current results do not necessarily imply that individual neurons are tuned to the mean or variability of a visual input; we do not wish to argue for “mean” or “variance” neurons. Instead, we suggest that the effects of input summary statistics on the decision process are compatible with a population representation of stimulus identity. Furthermore, the success of the LPR model in accounting for neural activity (both BOLD and EEG) suggests that neural activity does not necessarily direct reflect the presence or absence of particular sensory features, but instead reflects probability values. These probability values represent the probability of a particular feature being present, given the current input provided by the environment.

One approach to understanding neural coding proposes that instead of neural activity encoding a single best estimate of some stimulus value (parameter), populations of neurons actually encode a full probability distribution, allowing for a representation of uncertainty to be built into the neural code (Pouget, Beck, Ma, & Latham, 2013). Instead of neuronal activity reflecting a point estimate or best guess of the identity of a signal, it has been proposed that neuronal firing can also encode an estimate of the reliability of this estimate. An estimate of reliability is a very useful measure, since it indicates how much weight a signal should receive when guiding actions. Empirical studies have demonstrated that two independent sensory cues are weighted by their reliability when observers make sensory judgements, suggesting probabilistic information is readily encoded by a variety of neural networks (Ernst & Banks, 2002). The actual neural circuits underlying these probabilistic computations are, however, unclear (Ma & Jazayeri, 2014). For some, probabilistic information can be encoded by single neurons (Jazayeri & Movshon, 2006), but for others, a population code is assumed in which reliability is signalled by a

wider activation pattern across the population (Barber, Clark, & Anderson, 2003; Rao, 2004).

One specific prediction of the LPR model, arising from the resultant sigmoidal shape of the feature-decision space transfer function, is that the influence of some elements may be distorted. Specifically, elements lying towards the furthest points of feature space (in this case, the most prototypical colour or shapes) are underweighted considering their objective evidential value. This does not mean that these values are ignored, but it does mean that they are less influential than would be expected. One implementation of such a model could rely on different weightings given to neurons with particular tuning values, or it could assume a "hard-coded" weighting though neuronal populations of different sizes. Interestingly, there is a growing body of studies at the neuronal level which suggest that the distribution of tuning properties is not uniform across dimensions represented by neuronal populations, and that this scheme may facilitate the fine-tuning of neural representation (Buzsaki & Mizuseki, 2014) for demanding discriminations.

There are several neuronal coding schemes that could therefore account for the behavioural and neuroimaging findings presented here. Although there is considerable variation between these various coding schemes, it is important to note that none suggest explicit representation of summary values. Rather, these representations emerge from populations of neurons with different tuning preferences. The current work also suggests, however, that these summary values are not epiphenomenal; they do provide useful information for decision systems. This was particularly the case for the variability, which seems to be a cue for sensory adaptation.

7.4. Summary values as cues for adaptation.

One clear conclusion from Chapter 5 and 6 is that representation of summary variance is flexible, and is dependent on recent stimulus history. Furthermore, this flexibility in neuronal representation is associated with behavioural adaptation. In this case, response times were facilitated when the variability of the target matched a preceding prime array.

The same was not true for the mean distance to boundary, although there was a strong categorical effect, such that prime-target pairs drawn from the same category were associated with faster responses. Behavioural and neural priming is often used to make inferences about which features of a stimulus are encoded automatically (Grill-Spector et al., 2006), which in this case would again indicate that summary information is rapidly extracted from groups of stimuli.

The nature of the priming effect was highly reminiscent of so-called conflict adaptation. In this case, the repetition of a demanding trial (e.g. an incongruent Stroop stimulus, or a flanker task trial in which flanker and target are in opposing directions) leads to improved behavioural performance (Gratton et al., 1992). Although there has been some concern about whether such effects are due to partial feature repetitions (Hommel, 2004), the same behavioural facilitation for trials with congruent cognitive demands remains even when such low-level visual components are controlled. One prominent explanation of conflict adaptation sees this behavioural modulation as the end product of the conflict monitoring network associated with mPFC and dlPFC function (Botvinick et al., 2001; Kerns et al., 2004). This adaptation is therefore driven by a high-level modal monitoring system. This interpretation might, however, have predicted that we should observe a parallel adaptation to the mean in Chapter 5. However, adaptation was selective for the congruence of the prime-target variability.

There are several potential explanations for the lack of adaptation to the mean. Although this experiment failed to find such an effect, it could be that adaptation to the mean could emerge under different task regimes. For example, the longest prime-target interval was 500ms, and it could be that adaptation to the mean would only manifest after a longer period of time. Perhaps more likely, however, might be that a comparable adaptation might occur between trials i.e. after task-relevant stimuli. This might suggest that any automatic extraction of a summary average, as observed behaviourally, is not sufficient to drive adaptive processes. Instead, for the mean to be a cue for adaptation, it must be associated with a particular response option.

This latter explanation perhaps mirrors a distinction between conflict that can emerge at the motor (i.e. response) level and conflict at the perceptual level. It has been suggested that adaptation effects may even be driven primarily by stimulus conflict (Notebaert & Verguts, 2006). Several features of the behavioural data suggest that priming by variability is an effect that acts on early visual representations. Firstly, although there was also a weak priming effect on the irrelevant dimension (for Chapter 5, but not in Chapter 6), there was no significant between-dimension priming. This suggests that the adaptation was occurring at the level of the features, rather than on some abstract representation of variability. Adaption to the irrelevant dimension is interesting because it suggests that the noise on a task irrelevant dimension can influence the processing of the task-relevant dimension. Perhaps this could be interpreted as a form of distractor reduction or perceptual load reduction - the irrelevant dimension becomes less distracting when its distribution of features is less surprising (i.e. congruent variability). Secondly, the rapid time course might suggest reliance on an early stage in the processing hierarchy particularly since the effect was weakest at the longest prime-target time intervals.

This led to a prediction of the involvement of sensory-specific cortices in mediating the effect of priming by variability. Although several brain regions showed the predicted repetition suppression for repeated variance, the most consistent effects were in sensory cortices. In neural terms, these results are consistent with a scheme in which neural populations dynamically alter their sensitivity to best represent the statistics of the input (Barlow, 1961; Simoncelli & Olshausen, 2001; Wark et al., 2007). This suggests that there is no intrinsic cost to encountering a variable visual input, since the behavioural cost largely arises from a mis-match between the statistics of the current input and the representational settings of the relevant neural population.

7.5. Final Conclusions

This thesis has therefore demonstrated, across four experiments, that there are dissociable neural and behavioural effects of two sources of perceptual uncertainty when the decision requires the integration of more than one source of information: mean distance to the category boundary and the variability of the evidence under consideration. The effects of

mean and variance can be accounted for by a model in which observers integrate probability values that reflect the experienced likelihood of one response option given the current evidence. Although this model can account for independent behavioural effects with this single transformation into probabilistic decision space, the neural implementation of such computations may actually be distributed across the visual hierarchy. The influence of variability can be seen in the early stages of sensory representations, whilst the effects of mean are most keenly felt for representations of information in a decision-relevant space. Dissociating these two sources of perceptual uncertainty has therefore provided insight into the nature of information representation and transformation that underlies the most basic visual discriminations.

Appendices

Appendix 1

Figure A1.1

Eyetracking. Although participants were asked to fixate the central point throughout the task, in the fMRI experiment we collected eyetracking data from a subset of participants (n=10) to ensure that eye movements did not differ between conditions. Most participants held fixation well; the full trace from the best and worst participants are shown in Fig S4a. When we calculated the mean and standard deviation of the distance (path length) on each trial and compared these values across mean and variance levels (figure S4b) no differences in the mean or standard deviation of the trial eye path length were found for the relevant [all p-values>0.33] or irrelevant conditions [all p-values>0.19]. In the EEG experiment, trials with eye movements were removed offline.

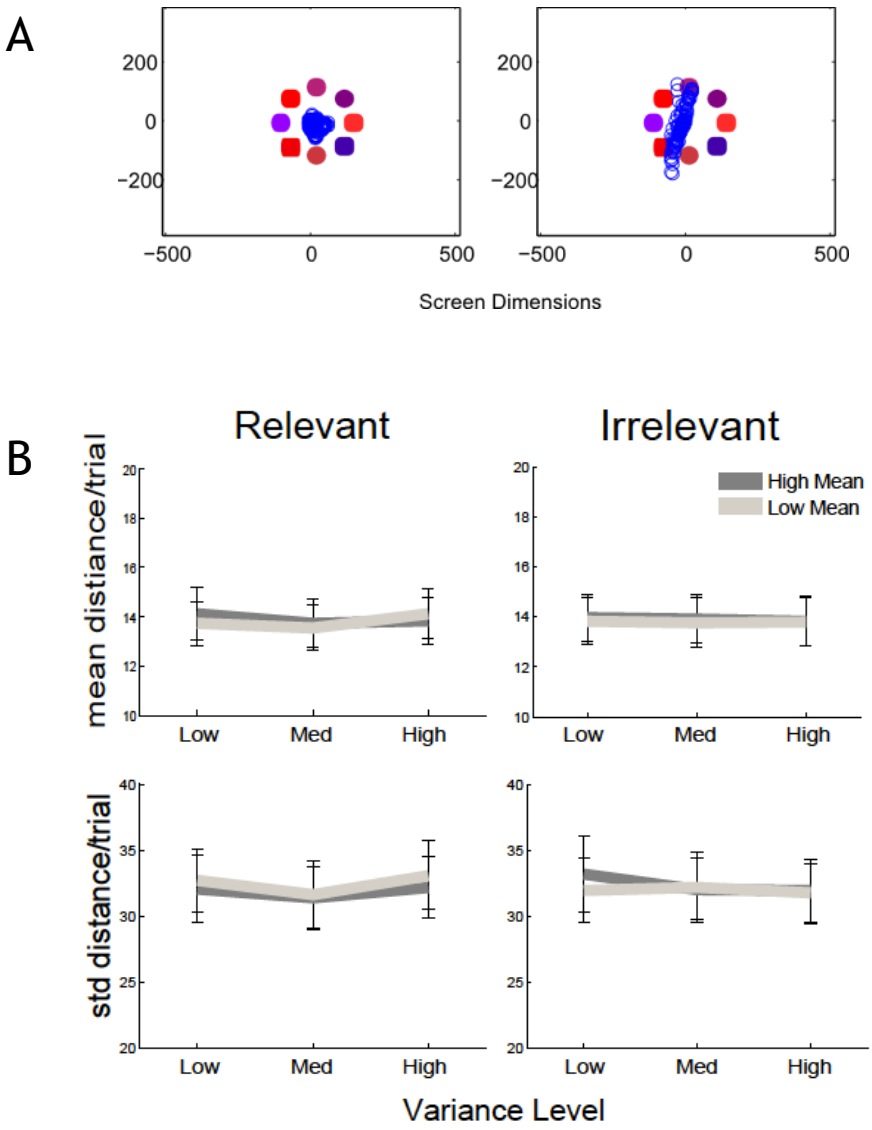


Figure A1.1 Eyetracking results. (**a**) Eye traces for the least (left) and most (right) variable participants for whom enough data was recorded. Each blue ring represents the coordinate at each sample time. An example array is presented at an approximate location for reference. (**b**) No difference was found in the mean eye path length (top row) or standard deviation in the eye path length (bottom row) for high vs low variance or high vs low mean. This was true for both the relevant (left column) and irrelevant (right column) dimensions.

Figure A1.2

Ventromedial PFC and posterior cingulate cortex. The ventromedial PFC (vmPFC) and posterior cingulate (PCC) tend to correlate positively with the value of the chosen option, the probability that a correct response will be selected – in other words, they are negatively activated by the presence of uncertainty. Here, we observed that the vmPFC (left peak: -2, 44, -2: $t_{(20)}$=5.13, p<0.0001; right peak: 2, 52, 2: $t_{(20)}$=4.80, p<0.0001) and PCC [-2, -52, 30 $t_{(20)}$=5.12, p<0.001] were negatively activated only by U_{Mr}. No correlation was observed with U_{Vr} [all p-values > 0.05]. BOLD signal in these regions showed no correlation with task-irrelevant mean or variance.

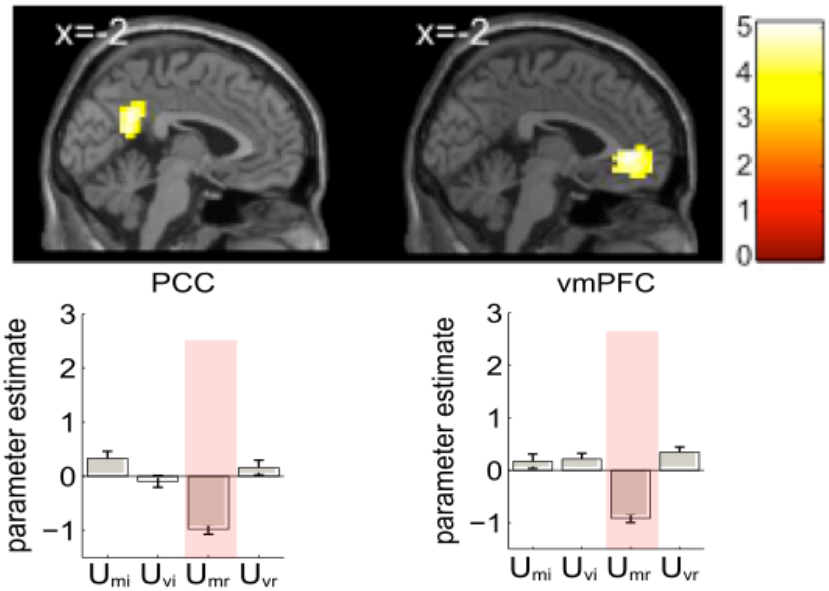


Figure A1.2. Imaging results in ventromedial prefrontal and posterior cingulate cortex. Both images are displayed with a threshold of P<0.001. A negative correlation with U_{Vr} was found in both the posterior cingulate cortex (PCC; **a**) and the ventromedial prefrontal cortex (vmPFC; **b**). The corresponding bar plots show that these regions responded only to $U_{Mr,}$ on the basis of which these regions were defined (red shading).

Table A1.1.1 Voxels Correlating with U_m for the Decision Space Analysis

STATISTICS: p-values adjusted for search volume

cluster p(FWE-cor)	cluster p(FDR-cor)	cluster equivk	cluster p(unc)	peak p(FWE-cor)	peak p(FDR-cor)	peak T	peak equivZ	peak p(unc)	x,y,z {mm}
0.000	0.000	126	0.000	0.069	0.153	6.22	4.59	0.000	-34 -48 50
				0.560	0.408	5.03	4.00	0.000	-50 -28 30
				0.999	0.778	3.79	3.25	0.001	-62 -20 26
0.000	0.000	290	0.000	0.071	0.153	6.20	4.58	0.000	34 24 -2
				0.497	0.385	5.14	4.05	0.000	54 16 14
				0.647	0.466	4.90	3.92	0.000	50 8 26
0.001	0.001	69	0.000	0.088	0.153	6.11	4.54	0.000	-30 28 -2
				0.095	0.153	6.07	4.52	0.000	-30 20 2
0.000	0.000	101	0.000	0.186	0.207	5.77	4.38	0.000	-46 4 26
				0.443	0.382	5.23	4.10	0.000	-54 4 38
				0.943	0.561	4.30	3.58	0.000	-58 12 10
0.004	0.002	57	0.000	0.450	0.382	5.22	4.10	0.000	10 24 42
0.459	0.177	12	0.064	0.773	0.553	4.69	3.81	0.000	-46 -60 -10
0.004	0.002	56	0.000	0.803	0.560	4.64	3.78	0.000	34 -64 46
				0.912	0.561	4.40	3.64	0.000	38 -44 54
0.115	0.045	24	0.013	0.912	0.561	4.40	3.64	0.000	-50 32 30
				0.979	0.596	4.12	3.47	0.000	-46 44 22
				0.992	0.667	3.98	3.37	0.000	-38 28 18
0.366	0.148	14	0.047	0.986	0.619	4.06	3.43	0.000	50 -32 46
				0.998	0.778	3.82	3.27	0.001	62 -20 42
0.042	0.017	33	0.005	0.024	0.078	6.70	4.80	0.000	66 -12 -14
				0.412	0.294	5.29	4.13	0.000	62 -4 -22
0.000	0.000	120	0.000	0.504	0.294	5.13	4.05	0.000	-2 44 -2
0.000	0.000	101	0.000	0.507	0.294	5.12	4.05	0.000	-2 -52 30
				0.610	0.314	4.95	3.95	0.000	-2 -56 18
				0.986	0.591	4.06	3.43	0.000	-6 -48 6
0.290	0.098	16	0.036	0.862	0.544	4.52	3.71	0.000	18 36 42

Voxels correlating with U_{mr} in the decision space analysis at a threshold of $p<0.001$ uncorrected for clusters larger than 10 voxels.

The following abbreviations have been used for the headings above and for all subsequent tables:

Cluster p(FWE-cor): clusterwise p-value with familywise-error correction for multiple comparisons; Cluster p(FDR-cor): clusterwise p-value with false discovery rate correction for multiple comparisons; Cluster equivk: number of voxels in cluster; cluster p(unc): uncorrected clusterwise p-value; voxel p(FWE): voxelwise p-value with family-wise error correction for multiple comparisons; voxel p(FDR): voxelwise p-value with false discovery rate correction for multiple comparisions; voxel T; voxelwise t-value; voxel equivZ: voxelwise z-score; voxel p(unc): voxelwise p-value, uncorrected; x,y,z {mm}: coordinated for the peak voxel, from the template brain of the Montreal Neurological Institute.

Table A1.1.2 Voxels Correlating with U_v for the Decision Space Analysis.

STATISTICS: p-values adjusted for search volume

cluster p(FWE-cor)	cluster p(FDR-cor)	cluster equivk	cluster p(unc)	peak p(FWE-cor)	peak p(FDR-cor)	peak T	peak equivZ	peak p(unc)	x,y,z {mm}
0.002	0.005	57	0.000	0.201	0.349	5.73	4.36	0.000	6 32 42
0.022	0.023	36	0.002	0.203	0.349	5.73	4.36	0.000	50 -28 -2
0.371	0.187	13	0.043	0.441	0.489	5.30	4.14	0.000	-10 -84 -30
0.195	0.111	18	0.020	0.740	0.847	4.81	3.88	0.000	26 56 -14
				0.991	0.920	4.07	3.43	0.000	26 64 -6
0.117	0.085	22	0.012	0.885	0.907	4.53	3.72	0.000	-2 -84 38
				0.994	0.920	4.01	3.40	0.000	-6 -96 26
				1.000	0.929	3.66	3.17	0.001	6 -80 30
0.421	0.187	12	0.051	0.973	0.920	4.23	3.53	0.000	38 56 18
				0.998	0.920	3.91	3.33	0.000	34 44 6
0.000	0.000	119	0.000	0.001	0.007	8.15	5.35	0.000	-30 -84 14
				0.012	0.013	7.04	4.94	0.000	-34 -84 6
0.000	0.000	331	0.000	0.006	0.011	7.37	5.07	0.000	34 -76 14
				0.138	0.053	5.90	4.44	0.000	26 -56 54
				0.152	0.053	5.86	4.42	0.000	22 -52 46
0.000	0.000	107	0.000	0.295	0.069	5.56	4.27	0.000	-18 -60 54

Voxels correlating with U_{vr} in the decision space analysis at a threshold of $p<0.001$ uncorrected for clusters larger than 10 voxels.

Table A1.1.3 Voxels Correlating with U_{mi} for the Decision Space Analysis.

STATISTICS: p-values adjusted for search volume

cluster p(FWE-cor)	cluster p(FDR-cor)	cluster equivk	cluster p(unc)	peak p(FWE-cor)	peak p(FDR-cor)	peak T	peak equivZ	peak p(unc)	x,y,z {mm}
0.445	0.273	13	0.068	0.944	0.995	4.23	3.54	0.000	18 -20 30
				0.953	0.995	4.20	3.51	0.000	10 -8 22
				0.996	0.995	3.85	3.29	0.001	22 -16 22

Voxels correlating with U_{mi} in the native space analysis at a threshold of $p<0.001$ uncorrected for clusters larger than 10 voxels. No voxels correlated significantly with U_{vi} at this threshold.

Table A1.2.1 Voxels Correlating Significantly with U_m in Native Space Analysis

STATISTICS: p-values adjusted for search volume

cluster p(FWE-cor)	cluster p(FDR-cor)	cluster equivk	cluster p(unc)	peak p(FWE-cor)	peak p(FDR-cor)	peak T	peak equivZ	peak p(unc)	x,y,z {mm}
0.000	0.000	155	0.000	0.154	0.333	5.85	4.42	0.000	-34 -44 46
				0.194	0.333	5.75	4.37	0.000	-38 -40 38
				0.542	0.364	5.08	4.02	0.000	-38 -52 62
0.000	0.000	146	0.000	0.285	0.333	5.56	4.28	0.000	38 -44 54
				0.335	0.333	5.45	4.22	0.000	30 -60 46
				0.500	0.364	5.14	4.06	0.000	22 -68 50
0.045	0.031	32	0.005	0.523	0.364	5.11	4.04	0.000	-30 16 6
0.198	0.086	19	0.022	0.713	0.509	4.80	3.87	0.000	-46 -64 -10
0.087	0.047	26	0.009	0.934	0.729	4.35	3.61	0.000	54 16 18
				0.996	0.771	3.90	3.33	0.000	46 8 26
0.223	0.086	18	0.026	0.965	0.730	4.22	3.53	0.000	-46 8 18
0.359	0.129	14	0.045	0.972	0.730	4.18	3.50	0.000	30 24 2
0.452	0.367	12	0.061	0.917	0.801	4.40	3.64	0.000	62 -8 -18
0.000	0.000	155	0.000	0.110	0.335	6.00	4.49	0.000	-34 -44 46
				0.171	0.335	5.80	4.39	0.000	-38 -40 38
				0.548	0.347	5.06	4.02	0.000	-38 -52 62
0.000	0.000	145	0.000	0.279	0.335	5.58	4.28	0.000	38 -44 54
				0.367	0.347	5.38	4.18	0.000	30 -60 46
				0.483	0.347	5.17	4.07	0.000	22 -68 50
0.040	0.025	33	0.004	0.523	0.347	5.10	4.04	0.000	-30 16 6
0.252	0.089	17	0.030	0.770	0.562	4.71	3.82	0.000	-46 -64 -10
0.157	0.079	21	0.017	0.943	0.635	4.31	3.58	0.000	54 16 18
				0.998	0.775	3.86	3.30	0.000	46 8 26
0.563	0.191	10	0.085	0.971	0.635	4.19	3.51	0.000	-2 -80 -34
				0.982	0.650	4.11	3.46	0.000	6 -76 -30
0.360	0.117	14	0.046	0.972	0.635	4.18	3.50	0.000	30 24 2
0.252	0.089	17	0.030	0.974	0.635	4.17	3.50	0.000	-46 8 18
0.453	0.369	12	0.062	0.935	0.835	4.34	3.60	0.000	62 -8 -18

Voxels correlating with U_{mr} in the native space analysis at a threshold of $p<0.001$ uncorrected for clusters larger than 10 voxels.

Table A1.2.2 Voxels Correlating with U_v in the Native Space Analysis

STATISTICS: p-values adjusted for search volume

cluster p(FWE-cor)	cluster p(FDR-cor)	cluster equivk	cluster p(unc)	peak p(FWE-cor)	peak p(FDR-cor)	peak T	peak equivZ	peak p(unc)	x,y,z {mm}
0.000	0.000	487	0.000	0.071	0.087	6.20	4.58	0.000	38 -84 22
				0.102	0.087	6.04	4.50	0.000	22 -68 42
				0.116	0.087	5.98	4.48	0.000	34 -76 14
0.000	0.000	417	0.000	0.198	0.087	5.74	4.36	0.000	-14 -64 50
				0.207	0.087	5.72	4.35	0.000	-30 -84 14
				0.314	0.103	5.51	4.25	0.000	-34 -80 -6
0.030	0.007	35	0.003	0.839	0.302	4.60	3.75	0.000	26 4 70
				0.943	0.406	4.33	3.59	0.000	38 0 62
				0.982	0.496	4.13	3.47	0.000	22 0 50
0.024	0.029	37	0.002	0.504	0.625	5.15	4.06	0.000	10 40 38
				0.904	0.625	4.45	3.67	0.000	10 32 42
				0.909	0.625	4.44	3.66	0.000	-6 44 30
0.555	0.221	10	0.080	0.530	0.625	5.11	4.04	0.000	2 20 66
0.497	0.214	11	0.068	0.709	0.625	4.83	3.88	0.000	2 -32 10
0.149	0.088	21	0.016	0.718	0.625	4.81	3.88	0.000	26 -44 14
				0.737	0.625	4.78	3.86	0.000	14 -36 18
0.027	0.029	36	0.003	0.753	0.625	4.75	3.84	0.000	46 -68 38
0.497	0.214	11	0.068	0.882	0.625	4.50	3.70	0.000	66 -44 10
0.215	0.106	18	0.024	0.929	0.625	4.38	3.62	0.000	-42 -68 46
0.058	0.044	29	0.006	0.955	0.640	4.28	3.57	0.000	2 -88 30
				0.988	0.664	4.07	3.43	0.000	6 -84 38
				0.993	0.664	4.00	3.39	0.000	-6 -92 30

Voxels correlating with U_{vr} in the native space analysis at a threshold of $p<0.001$ uncorrected for clusters larger than 10 voxels.

Table A1.2.3 Voxels Correlating with U_{vi} in the Native Space Analysis.

STATISTICS: p-values adjusted for search volume

cluster p(FWE-cor)	cluster p(FDR-cor)	cluster equivk	cluster p(unc)	peak p(FWE-cor)	peak p(FDR-cor)	peak T	peak equivZ	peak p(unc)	x,y,z {mm}
0.530	0.269	11	0.084	0.490	0.698	5.11	4.04	0.000	70 0 18
				0.594	0.698	4.94	3.95	0.000	66 4 26
0.585	0.269	10	0.098	0.903	0.698	4.39	3.63	0.000	-26 -76 -10
0.023	0.029	41	0.003	0.927	0.698	4.32	3.59	0.000	-26 -84 26
				0.947	0.698	4.25	3.54	0.000	-26 -88 34
0.105	0.068	26	0.012	0.981	0.714	4.07	3.43	0.000	38 -84 18
				0.982	0.714	4.06	3.42	0.000	50 -76 22
				0.996	0.823	3.87	3.31	0.000	30 -76 22

Voxels correlating with U_{vi} in the native space analysis at a threshold of $p<0.001$ uncorrected for clusters larger than 10 voxels. No voxels were found to correlate with U_{mi} at this threshold.

Appendix 2

Table A2.1.

ANOVA results. Factors: relevant dimension mean (mu: low/high), relevant dimension variance (sig: low/med/high), irrelevant dimension mean (irrmu: low/high), irrelevant dimension variance (irrsig: low/med/high). Analysis conducted on RTs, only correct trials included.

VARIABLE	DoF	F-VALUE	P-VALUE
Mu	1,18	106	<0.001 ***
Sig	2, 30	24.4	<0.001 ***
Mu x Sig	2, 34	0.64	0.525
Irrmu	1, 18	1.74	0.204
Mu x Irrmu	1, 18	1.11	0.305
Sig x Irrmu	2, 32	0.23	0.773
Mu x Sig x Irrmu	2, 28	0.50	0.569
Irrsig	2, 32	4.74	0.019 *
Mu x Irrsig	2, 34	0.56	0.566
Sig x Irrsig	3, 53	2.79	0.051
Mu x Sig x Irrsig	3, 50	0.93	0.428
Irrmu x Irrsig	2, 34	3.43	0.046 *
Mu x Irrmu x Irrsig	2, 32	1.87	0.174
Sig x Irrmu x Irrsig	3, 50	0.69	0.551
Mu x Sig x Irrmu x Irrsig	3, 54	1.53	0.217

Table A2.2.

ANOVA results. Factors: relevant dimension mean (mu: low/high), relevant dimension variance (sig: low/med/high), irrelevant dimension mean (irrmu: low/high), irrelevant dimension variance (irrsig: low/med/high). Analysis conducted on error rates.

VARIABLE	DoF	F-VALUE	P-VALUE
Mu	1, 18	66.3	<0.001 ***
Sig	1.71, 30.9	9.93	0.001 ***
Mu x Sig	1.98, 35.6	1.38	0.264
Irrmu	1, 18	0.065	0.802
Mu x Irrmu	1,18	4.95	0.039 *
Sig x Irrmu	1.97, 35.4	0.25	0.779
Mu x Sig x Irrmu	1.54, 27.6	5.82	0.012 *
Irrsig	1.78, 32.0	0.51	0.584
Mu x Irrsig	1.82, 32.8	1.66	0.207
Sig x Irrsig	3.38, 60.93	1.13	0.348
Mu x Sig x Irrsig	2.90, 52.1	2.54	0.068
Irrmu x Irrsig	1.74, 31.3	0.11	0.874
Mu x Irrmu x Irrsig	1.74, 31.4	0.73	0.474
Sig x Irrmu x Irrsig	3.07, 55.2	4.78	0.005 **
Mu x Sig x Irrmu x Irrsig	3.25, 58.6	0.69	0.573

Appendix 3

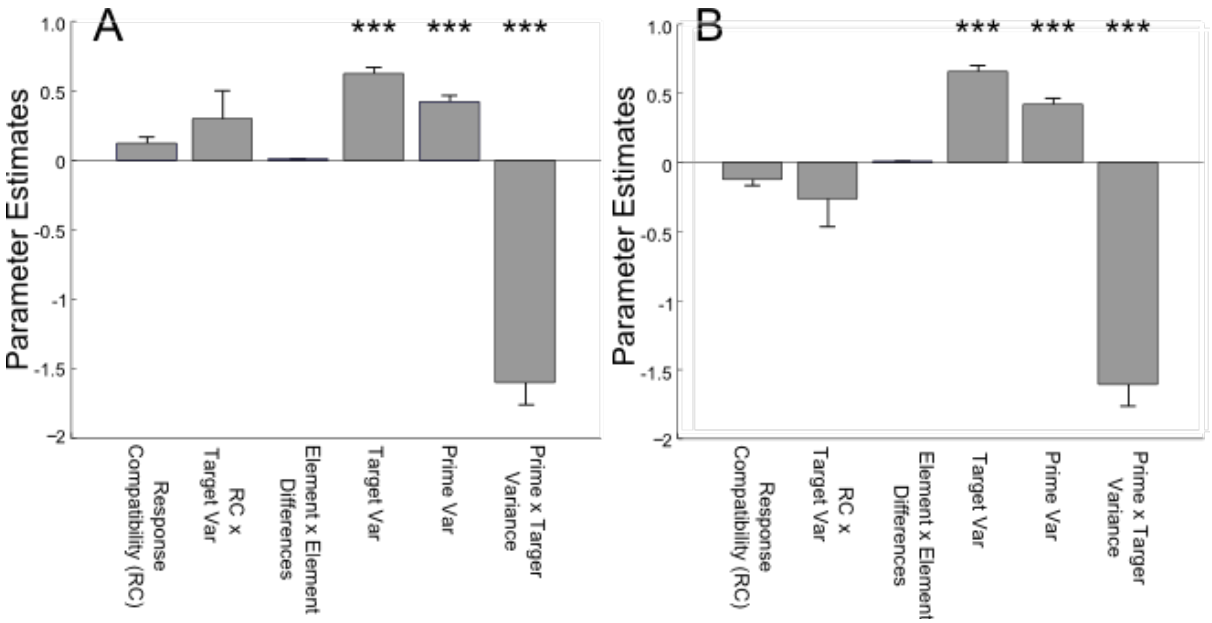


Fig A3.1. Parameter estimates from a regression of prime and target statistics on response times, with additional response compatibility regressors included. Data from Experiment 1. All error bars show standard error of the mean. **(A)** Parameter estimates for (i) response compatibility measure (RC), defined as the absolute difference between the sum of feature values for prime and target arrays ($|\Sigma P_{1\text{-}8}\text{-}T_{1\text{-}8}|$), and (ii) its interaction with target variance, alongside (iii) the summed element by element difference (as shown individually in Fig 2B), (iv) target variance, (v) prime variance and (vi) the interaction between prime and target variance. **(B)** As for (A), with RC defined instead as the sum of the prime and target feature values ($|\Sigma P_{1\text{-}8}+T_{1\text{-}8}|$).

Tables A3.1

Effect of Prime and Target Statistics

Table A3.1.1 Experiment 1. ANOVA with 2 factors (prime variance (Pvar) or target variance (Tvar), each with 2 levels (high/low). Stars indicate level of significance as follows for all tables: $p < 0.05$ (*), $p < 0.01$ (**) or $p < 0.001$ (***).

VARIABLE	DOF	F – VALUE	P- VALUE
Tvar	1, 39	76.9	<0.001 ***
Pvar	1, 39	5.24	0.028 *
Tvar*Pvar	1, 39	27.5	<0.001 ***

Table A3.1.2 Experiment 2. ANOVA with 2 factors (prime/target variance), each with 2 levels (high/low).

VARIABLE	DOF	F – VALUE	P- VALUE
Tvar	1, 38	37.8	<0.001 ***
Pvar	1, 38	1.28	0.267
Tvar*Pvar	1, 38	5.03	0.031 *

Table **A3.1.3 Experiment 2.** ANOVA with 2 factors (prime/target mean), each with 2 levels (high/low).

VARIABLE	DOF	F – VALUE	P- VALUE
Tmean	1, 38	63.6	<0.001 ***
Pmean	1, 38	0.04	0.841
Tmean*Pmean	1, 38	0.20	0.661

Table A3.2

Effect of PTI

Table **A3.2 Experiments 1 and 2.** ANOVA with 3 factors: prime/target variance and PTI duration. Prime and target variance had 2 levels (high/low) and PTI had 3 levels (100, 200 or 500ms).

VARIABLE	DOF	F – VALUE	P - VALUE
Tvar	1,78	105	<0.001 ***
Pvar	1,78	6.37	0.014 *
Pvar*Tvar	1,78	26.2	<0.001 ***
PTI	2,121	16.7	<0.001 ***
PTI*Tvar	2,153	1.31	0.273
PTI*Pvar	2,155	0.80	0.450
PTI*Pvar*Tvar	2,147	1.97	0.146

Tables A3.3

Effect of Category Congruence on RTs - 100 and 200ms Conditions

Table A3.3.1 Experiment 1. ANOVA with 3 factors (prime/target variance and category congruence – "switch"), each with 2 levels (high/low or switch/stay).

VARIABLE	DOF	F – VALUE	P - VALUE
Tvar	1,39	67.1	<0.001 ***
Switch	1,39	32.3	<0.001 ***
Tvar*Switch	1,39	1.23	0.275
Pvar	1,39	2.63	0.113
Pvar*Switch	1,39	13.6	<0.001 ***
Pvar*Tvar	1,39	30.1	<0.001 ***
Pvar*Tvar*Switch	1,39	5.74	0.022 *

Table A3.3.2 Experiment 2. ANOVA with 3 factors (prime/target variance and category congruence – "switch"), each with 2 levels (high/low or switch/stay).

VARIABLE	DOF	F – VALUE	P - VALUE
Tvar	1,38	32.4	<0.001 ***
Switch	1,38	24.6	<0.001 ***
Tvar*Switch	1,38	0.01	0.993
Pvar	1,38	1.07	0.308
Pvar*Switch	1,38	2.18	0.148
Pvar*Tvar	1,38	4.59	0.039
Pvar*Tvar*Switch	1,38	9.05	0.005 **

Table A3.3.3 Experiment 3. ANOVA with 3 factors (prime/target variance and category congruence – "switch"), each with 2 levels (high/low or switch/stay).

VARIABLE	DOF	F – VALUE	P - VALUE
Tvar	1,13	18.9	0.001***
Switch	1,13	2.52	0.137
Tvar*Switch	1,13	1.98	0.183
Pvar	1,13	8.64	0.011
Pvar*Switch	1,13	2.52	0.136
Pvar*Tvar	1,13	13.7	0.003 **
Pvar*Tvar*Switch	1,13	3.90	0.070

Tables A3.4

Effect of Category Congruence on RTs - all conditions (experiment 3 omitted here as it is identical to Table A3.3.3)

Table A3.4.1. Experiment 1. ANOVA with 3 factors (prime/target variance and category congruence – "switch"), each with 2 levels (high/low or switch/stay).

VARIABLE	DOF	F – VALUE	P - VALUE
Tvar	1,39	78.3	<0.001 ***
Switch	1,39	23.0	<0.001 ***
Tvar*Switch	1,39	0.48	0.492
Pvar	1,39	4.82	0.034 *
Pvar*Switch	1,39	16.5	<0.001 ***
Pvar*Tvar	1,39	27.1	<0.001 ***
Pvar*Tvar*Switch	1,39	9.42	0.004 **

Table A3.4.2. Experiment 2. ANOVA with 3 factors (prime/target variance and category congruence – "switch"), each with 2 levels (high/low or switch/stay).

VARIABLE	DOF	F – VALUE	P - VALUE
Tvar	1,38	38.3	<0.001 ***
Switch	1,38	20.9	<0.001 ***
Tvar*Switch	1,38	0.02	0.893
Pvar	1,38	0.99	0.326
Pvar*Switch	1,38	6.81	0.013 *
Pvar*Tvar	1,38	5.03	0.031
Pvar*Tvar*Switch	1,38	11.2	0.002 **

Table A3.5

Regression Analyses for Prime x Target Variance Interaction, for switch and stay trials. data collapsed across all PTI conditions.

	Switch	Stay
Overall	$t_{(92)} = 1.38$, $p < 0.09$	$t_{(92)} = 7.66$, $p < 0.001$
Experiment 1	$t_{(39)} = 1.44$, $p < 0.08$	$t_{(39)} = 6.15$, $p < 0.001$
Experiment 2	$t_{(38)} = 1.22$, $p < 0.884$	$t_{(38)} = 3.54$, $p < 0.001$
Experiment 3	$t_{(13)} = 2.30$, $p < 0.019$	$t_{(13)} = 3.91$, $p < 0.001$

Appendix 4

Table A4.1.1

Voxels Correlating with Target Variability.

cluster p(FWE-cor)	cluster p(FDR-cor)	cluster equivk	cluster p(unc)	peak p(FWE-cor)	peak p(FDR-cor)	peak T	peak equivZ	peak p(unc)	x,y,z {mm}
0.000	0.000	360	0.000	0.021	0.125	6.94	4.84	0.000	-34 -80 6
				0.053	0.128	6.48	4.65	0.000	-30 -76 14
				0.141	0.128	6.01	4.44	0.000	-38 -60 -2
0.000	0.000	254	0.000	0.126	0.128	6.06	4.47	0.000	26 -64 58
				0.294	0.128	5.66	4.28	0.000	46 -28 46
				0.472	0.204	5.29	4.10	0.000	34 -72 14
0.083	0.044	27	0.009	0.541	0.212	5.17	4.04	0.000	46 0 26
				0.727	0.248	4.87	3.87	0.000	58 8 18
0.131	0.044	23	0.014	0.715	0.248	4.89	3.88	0.000	-22 0 54
0.131	0.044	23	0.014	0.869	0.364	4.59	3.72	0.000	30 -4 46
0.209	0.062	19	0.023	0.965	0.479	4.29	3.54	0.000	30 -72 -6
				0.982	0.550	4.17	3.47	0.000	38 -60 -10

Data shown for clusters larger than 10 voxels for this and all subsequent tables.

The following abbreviations have been used for the headings above and for all subsequent tables:

Cluster p(FWE-cor): clusterwise p-value with familywise-error correction for multiple comparisons; Cluster p(FDR-cor): clusterwise p-value with false discovery rate correction for multiple comparisons; Cluster equivk: number of voxels in cluster; cluster p(unc): uncorrected clusterwise p-value; voxel p(FWE): voxelwise p-value with family-wise error correction for multiple comparisons; voxel p(FDR): voxelwise p-value with false discovery rate correction for multiple comparisions; voxel T; voxelwise t-value; voxel equivZ: voxelwise z-score; voxel p(unc): voxelwise p-value, uncorrected; x,y,z {mm}: coordinated for the peak voxel, from the template brain of the Montreal Neurological Institute.

Table A4.1.2

Voxels Correlating with Prime-Target Variance Interaction Term.

cluster p(FWE-cor)	cluster p(FDR-cor)	cluster	cluster p(unc)	peak equivk	peak p(unc)	peak p(FWE-cor)	peak	peak p(FDR-cor)	T	equivZ	x,y,z {mm}
0.070	0.043	34	0.009	0.366	0.767	5.35	4.13	0.000	-54	-48	-10
0.077	0.043	33	0.010	0.561	0.767	5.00	3.95	0.000	26	-8	54
0.032	0.043	43	0.004	0.731	0.802	4.72	3.79	0.000	34	24	18
				0.822	0.802	4.55	3.70	0.000	34	4	30
				0.998	0.931	3.76	3.21	0.001	46	4	34

Table A4.2.1

Voxels Correlating with Irrelevant Target Variance.

cluster p(FWE-cor)	cluster p(FDR-cor)	cluster	cluster p(unc)	peak equivk	peak p(unc)	peak p(FWE-cor)	peak	peak p(FDR-cor)	T		equivZ	peak x,y,z {mm}
0.030	0.032	42	0.004	0.640	0.989	4.90	3.89	0.000	-26	-76	2	
				0.993	0.989	3.93	3.32	0.000	-26	-76	14	
				1.000	0.989	3.62	3.12	0.001	-34	-76	22	
0.039	0.032	39	0.005	0.776	0.989	4.67	3.77	0.000	30	-68	14	
				0.899	0.989	4.41	3.62	0.000	34	-68	-2	
				0.989	0.989	3.99	3.36	0.000	46	-60	-6	
0.047	0.032	37	0.006	0.910	0.989	4.38	3.60	0.000	38	28	18	
				0.971	0.989	4.14	3.45	0.000	46	28	22	
				0.973	0.989	4.13	3.45	0.000	38	12	30	

Table A4.3

Voxels Correlating with EED.

STATISTICS: p-values adjusted for search volume

cluster p(FWE-cor)	cluster	cluster p(FDR-cor)	cluster	peak equivk	peak p(unc)	peak p(FWE-cor)	peak	peak p(FDR-cor)	T	equivZ	p(unc)	x,y,z {mm}
0.057	0.160	36	0.007	0.427	0.884	5.24	4.07	0.000	-54	-48	-10	
				0.455	0.884	5.19	4.05	0.000	-62	-48	-6	
0.336	0.288	17	0.050	0.665	0.884	4.83	3.86	0.000	34	24	18	
0.107	0.160	29	0.014	0.726	0.884	4.73	3.80	0.000	26	-8	54	
0.156	0.160	25	0.021	0.763	0.884	4.67	3.76	0.000	34	4	30	
				0.998	0.998	3.74	3.19	0.001	46	4	34	

References

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. Nat Neurosci, 14(10), 1338-1344. doi: 10.1038/nn.2921

Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. Annu Rev Neurosci, 8, 407-430. doi: 10.1146/annurev.ne.08.030185.002203

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. Trends Cogn Sci, 15(3), 122-131. doi: 10.1016/j.tics.2011.01.003

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. Psychological Science, 19(4), 392-398. doi: 10.1111/j.1467-9280.2008.02098.x

Andersen, R. A., & Buneo, C. A. (2002). Intentional maps in posterior parietal cortex. Annu Rev Neurosci, 25, 189-220. doi: 10.1146/annurev.neuro.25.112701.142922

Ariely, D. (2001). Seeing sets: Representation by statistical properties. Psychological Science, 12(2), 157-162. doi: Doi 10.1111/1467-9280.00327

Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? Percept Psychophys, 70(7), 1325-1326; discussion 1335-1326. doi: 10.3758/PP.70.7.1325

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What Some Concepts Might Not Be. Cognition, 13(3), 263-308. doi: Doi 10.1016/0010-0277(83)90012-4

Ashby, F. G., & Gott, R. E. (1988). Decision Rules in the Perception and Categorization of Multidimensional Stimuli. Journal of Experimental Psychology-Learning Memory and Cognition, 14(1), 33-53. doi: Doi 10.1037/0278-7393.14.1.33

Attneave, F. (1954). Some informational aspects of visual perception. Psychol Rev, 61(3), 183-193.

Baker, C. I., Behrmann, M., & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nat Neurosci, 5(11), 1210-1216. doi: 10.1038/nn960

Baker, J. F., Petersen, S. E., Newsome, W. T., & Allman, J. M. (1981). Visual response properties of neurons in four extrastriate visual areas of the owl monkey (Aotus trivirgatus): a quantitative comparison of medial, dorsomedial, dorsolateral, and middle temporal areas. J Neurophysiol, 45(3), 397-416.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. J Vis, 9(12), 13 11-18. doi: 10.1167/9.12.13

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. J Cogn Neurosci, 7(1), 66-80. doi: 10.1162/jocn.1995.7.1.66

Barber, M. J., Clark, J. W., & Anderson, C. H. (2003). Neural representation of probabilistic information. Neural Comput, 15(8), 1843-1864. doi: 10.1162/08997660360675062

Barlow, H. B. (1956). Retinal noise and absolute threshold. J Opt Soc Am, 46(8), 634-639.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), Sensory Communication (pp. 217-234). Cambridge, MA: MIT Press.

Barlow, H. B., Levick, W. R., & Yoon, M. (1971). Responses to single quanta of light in retinal ganglion cells of the cat. Vision Res, Suppl 3, 87-101.

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. Proc Natl Acad Sci U S A, 107(50), 21767-21772. doi: 10.1073/pnas.0908104107

Baylis, G. C., & Rolls, E. T. (1987). Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. Exp Brain Res, 65(3), 614-622.

Beason-Held, L. L., Purpura, K. P., Krasuski, J. S., Maisog, J. M., Daly, E. M., Mangot, D. J., . . . VanMeter, J. W. (1998). Cortical regions involved in visual texture perception: a fMRI study. Brain Res Cogn Brain Res, 7(2), 111-118.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. Nat Neurosci, 10(9), 1214-1221. doi: 10.1038/nn1954

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. Vision Res, 37(23), 3327-3338.

Bennur, S., & Gold, J. I. (2011). Distinct representations of a perceptual decision and the associated oculomotor plan in the monkey lateral intraparietal area. J Neurosci, 31(3), 913-921. doi: 10.1523/JNEUROSCI.4417-10.2011

Biederman, I. (1972). Perceiving real-world scenes. Science, 177(4043), 77-80.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W., Jr. (1974). On the information extracted from a glance at a scene. Journal of Experimental Psychology, 103(3), 597-600.

Blair, M., & Homa, D. (2003). As easy to memorize as they are to classify: The 5-4 categories and the category advantage. Mem Cognit, 31(8), 1293-1301. doi: Doi 10.3758/Bf03195812

Blanchard, T. C., & Hayden, B. Y. (2014). Neurons in dorsal anterior cingulate cortex signal postdecisional variables in a foraging task. J Neurosci, 34(2), 646-655. doi: 10.1523/JNEUROSCI.3151-13.2014

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. Psychol Rev, 113(4), 700-765. doi: Doi 10.1037/0033-295x.113.4.700

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. Psychol Rev, 108(3), 624-652.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn Sci, 8(12), 539-546. doi: 10.1016/j.tics.2004.10.003

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. Vis Neurosci, 13(1), 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. J Neurosci, 12(12), 4745-4765.

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. Science, 307(5712), 1118-1121. doi: 10.1126/science.1105783

Brown, M. W., Wilson, F. A., & Riches, I. P. (1987). Neuronal evidence that inferomedial temporal cortex is more important than hippocampus in certain processes underlying recognition memory. Brain Res, 409(1), 158-162.

Buzsaki, G., & Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. Nat Rev Neurosci, 15(4), 264-278. doi: 10.1038/nrn3687

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. J Neurosci, 32(22), 7685-7700. doi: 10.1523/JNEUROSCI.3325-11.2012

Carandini, M. (2006). What simple and complex cells compute. J Physiol, 577(Pt 2), 463-466. doi: 10.1113/jphysiol.2006.118976

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. Science, 280(5364), 747-749.

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: an update of theory and data. Cogn Affect Behav Neurosci, 7(4), 367-379.

Chong, S. C., Joo, S. J., Emmanouil, T. A., & Treisman, A. (2008). Statistical processing: not so implausible after all. Percept Psychophys, 70(7), 1327-1334; discussion 1335-1326. doi: 10.3758/PP.70.7.1327

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. Vision Res, 43(4), 393-404. doi: Pii S0042-6989(02)00596-5 Doi 10.1016/S0042-6989(02)00596-5

Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. Percept Psychophys, 67(1), 1-13.

Cohen, Y. E. (2009). Multimodal activity in the parietal cortex. Hear Res, 258(1-2), 100-105. doi: 10.1016/j.heares.2009.01.011

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. Annu Rev Neurosci, 22, 319-349. doi: 10.1146/annurev.neuro.22.1.319

Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L., & King, A. J. (2010). Adaptation to stimulus statistics in the perception and neural representation of auditory space. Neuron, 66(6), 937-948. doi: 10.1016/j.neuron.2010.05.018

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. Proc Natl Acad Sci U S A, 108(32), 13341-13346. doi: 10.1073/pnas.1104517108

de Lange, F. P., Jensen, O., & Dehaene, S. (2010). Accumulation of evidence during sequential decision making: the importance of top-down factors. J Neurosci, 30(2), 731-738. doi: 10.1523/JNEUROSCI.4080-09.2010

de Lange, F. P., Rahnev, D. A., Donner, T. H., & Lau, H. (2013). Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. J Neurosci, 33(4), 1400-1410. doi: 10.1523/JNEUROSCI.1094-12.2013

Demeyere, N., Rzeskiewicz, A., Humphreys, K. A., & Humphreys, G. W. (2008). Automatic statistical processing of visual properties in simultanagnosia. Neuropsychologia, 46(11), 2861-2864. doi: 10.1016/j.neuropsychologia.2008.05.014

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. Proc Natl Acad Sci U S A, 93(24), 13494-13499.

Ding, L., & Gold, J. I. (2010). Caudate Encodes Multiple Computations for Perceptual Decisions. Journal of Neuroscience, 30(47), 15747-15759. doi: Doi 10.1523/Jneurosci.2894-10.2010

Ding, L., & Gold, J. I. (2012). Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. Cereb Cortex, 22(5), 1052-1067. doi: 10.1093/cercor/bhr178

Donner, T. H., Siegel, M., Fries, P., & Engel, A. K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. Curr Biol, 19(18), 1581-1585. doi: 10.1016/j.cub.2009.07.066

Dubner, R., & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. Brain Res, 35(2), 528-532.

Duffy, F. H., & Burchfiel, J. L. (1971). Somatosensory system: organizational hierarchy from single units in monkey area 5. Science, 172(3980), 273-275.

Ebbinghaus, H., & Dürr Ernst. (1902). Grundzüge der Psychologie. Leipzig: Viet.

Egner, T. (2007). Congruency sequence effects and cognitive control. Cogn Affect Behav Neurosci, 7(4), 380-390.

Erickson, C. A., & Desimone, R. (1999). Responses of macaque perirhinal neurons during and after visual stimulus association learning. J Neurosci, 19(23), 10404-10416.

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

Ewbank, M. P., Schluppeck, D., & Andrews, T. J. (2005). fMR-adaptation reveals a distributed representation of inanimate objects and places in human visual cortex. Neuroimage, 28(1), 268-279. doi: 10.1016/j.neuroimage.2005.06.036

Fairhall, A. L., Lewen, G. D., Bialek, W., & de Ruyter Van Steveninck, R. R. (2001). Efficiency and ambiguity in an adaptive neural code. Nature, 412(6849), 787-792. doi: 10.1038/35090500

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? J Vis, 7(1), 10. doi: 10.1167/7.1.10

Fitzgerald, J. K., Swaminathan, S. K., & Freedman, D. J. (2012). Visual categorization and the parietal cortex. Front Integr Neurosci, 6, 18. doi: 10.3389/fnint.2012.00018

Foldiak, P., Xiao, D., Keysers, C., Edwards, R., & Perrett, D. I. (2004). Rapid serial visual presentation for the determination of neural selectivity in area STSa. Prog Brain Res, 144, 107-116.

Frank, M. J., Samanta, J., Moustafa, A. A., & Sherman, S. J. (2007). Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science, 318(5854), 1309-1312. doi: 10.1126/science.1146157

Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. Nature, 443(7107), 85-88. doi: 10.1038/nature05078

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. Science, 291(5502), 312-316. doi: 10.1126/science.291.5502.312

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J Neurosci, 23(12), 5235-5246.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. Nat Neurosci, 14(9), 1195-1201. doi: 10.1038/nn.2889

Gepshtein, S., Lesmes, L. A., & Albright, T. D. (2013). Sensory adaptation as optimal resource allocation. Proc Natl Acad Sci U S A, 110(11), 4368-4373. doi: 10.1073/pnas.1204109110

Gerken, L., Dawson, C., Chatila, R., & Tenenbaum, J. (2015). Surprise! Infants consider possible bases of generalization for a single input example. Dev Sci, 18(1), 80-89. doi: 10.1111/desc.12183

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. Annu Rev Neurosci, 30, 535-574. doi: 10.1146/annurev.neuro.29.051605.113038

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. Trends Neurosci, 15(1), 20-25.

Gorea, A. (1995). Early vision and beyond. In T. V. Papathomas & C. Chubb (Eds.), (pp. xi, 272 p.). Cambridge, Mass ; London: MIT Press.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. Journal of Experimental Psychology-General, 121(4), 480-506.

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. New York ; London: Wiley.

Greene, M. R., & Oliva, A. (2009). The briefest of glances: the time course of natural scene understanding. Psychological Science, 20(4), 464-472.

Greene, M. R., & Oliva, A. (2010). High-level aftereffects to global scene properties. J Exp Psychol Hum Percept Perform, 36(6), 1430-1442. doi: 10.1037/a0019058

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci, 10(1), 14-23. doi: 10.1016/j.tics.2005.11.006

Grinband, J., Hirsch, J., & Ferrera, V. P. (2006). A neural representation of categorization uncertainty in the human brain. Neuron, 49(5), 757-763. doi: 10.1016/j.neuron.2006.01.032

Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. Neuroimage, 57(2), 303-311. doi: 10.1016/j.neuroimage.2010.12.027

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science, 166(3910), 1303-1306.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. Curr Biol, 17(17), R751-753. doi: 10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. J Exp Psychol Hum Percept Perform, 35(3), 718-734. doi: 10.1037/a0013899

Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. Psychon Bull Rev, 18(5), 855-859. doi: Doi 10.3758/S13423-011-0125-6

Haefner, R. M., Gerwinn, S., Macke, J. H., & Bethge, M. (2013). Inferring decoding strategies from choice probabilities in the presence of correlated variability. Nat Neurosci, 16(2), 235-242. doi: 10.1038/nn.3309

Hanks, T. D., Ditterich, J., & Shadlen, M. N. (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. Nat Neurosci, 9(5), 682-689. doi: 10.1038/nn1683

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science, 293(5539), 2425-2430. doi: 10.1126/science.1063736

Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. Nature, 431(7010), 859-862. doi: 10.1038/nature02966

Heekeren, H. R., Marrett, S., Ruff, D. A., Bandettini, P. A., & Ungerleider, L. G. (2006). Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. Proc Natl Acad Sci U S A, 103(26), 10023-10028. doi: 10.1073/pnas.0603949103

Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. Nat Rev Neurosci, 9(6), 467-479. doi: 10.1038/nrn2374

Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. Science, 287(5456), 1269-1272.

Ho, T. C., Brown, S., & Serences, J. T. (2009). Domain general mechanisms of perceptual decision making in human cortex. J Neurosci, 29(27), 8675-8687. doi: 10.1523/JNEUROSCI.5984-08.2009

Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. Neuron, 36(5), 791-804.

Hommel, B. (2004). Event files: feature binding in and across perception and action. Trends Cogn Sci, 8(11), 494-500. doi: 10.1016/j.tics.2004.08.007

Horwitz, G. D., Batista, A. P., & Newsome, W. T. (2004). Representation of an abstract perceptual decision in macaque superior colliculus. J Neurophysiol, 91(5), 2281-2296. doi: 10.1152/jn.00872.2003

Howe, C. Q., & Purves, D. (2004). Size contrast and assimilation explained by the statistics of natural scene geometry. J Cogn Neurosci, 16(1), 90-102. doi: 10.1162/089892904322755584

Howe, C. Q., & Purves, D. (2005). The Muller-Lyer illusion explained by the statistics of image-source relationships. Proc Natl Acad Sci U S A, 102(4), 1234-1239. doi: 10.1073/pnas.0409314102

Howe, C. Q., Yang, Z., & Purves, D. (2005). The Poggendorff illusion explained by natural scene geometry. Proc Natl Acad Sci U S A, 102(21), 7707-7712. doi: 10.1073/pnas.0502893102

Hubel, D. H. (1959). Single unit activity in striate cortex of unrestrained cats. J Physiol, 147, 226-238.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. J Physiol, 148, 574-591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol, 160, 106-154.

Huettel, S. A., Song, A. W., & McCarthy, G. (2005). Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. J Neurosci, 25(13), 3304-3311. doi: 10.1523/JNEUROSCI.5070-04.2005

Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. J Neurosci, 25(45), 10420-10436. doi: 10.1523/JNEUROSCI.4684-04.2005

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. Science, 310(5749), 863-866. doi: 10.1126/science.1117593

Ide, J. S., Shenoy, P., Yu, A. J., & Li, C. S. (2013). Bayesian prediction and evaluation in the anterior cingulate cortex. J Neurosci, 33(5), 2039-2047. doi: 10.1523/JNEUROSCI.2201-12.2013

Iwamura, Y. (1998). Hierarchical somatosensory processing. Curr Opin Neurobiol, 8(4), 522-528.

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. Nat Neurosci, 9(5), 690-696. doi: 10.1038/nn1691

Jocham, G., Hunt, L. T., Near, J., & Behrens, T. E. (2012). A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. Nat Neurosci, 15(7), 960-961. doi: 10.1038/nn.3140

Kato, R., Takaura, K., Ikeda, T., Yoshida, M., & Isa, T. (2011). Contribution of the retino-tectal pathway to visually guided saccades after lesion of the primary visual cortex in monkeys. Eur J Neurosci, 33(11), 1952-1960. doi: 10.1111/j.1460-9568.2011.07729.x

Kayaert, G., Biederman, I., & Vogels, R. (2005). Representation of regular and irregular shapes in macaque inferotemporal cortex. Cereb Cortex, 15(9), 1308-1321. doi: 10.1093/cercor/bhi014

Kayser, A. S., Buchsbaum, B. R., Erickson, D. T., & D'Esposito, M. (2010). The functional anatomy of a perceptual decision in the human brain. J Neurophysiol, 103(3), 1179-1194. doi: 10.1152/jn.00364.2009

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. J Neurosci, 33(50), 19434-19441. doi: 10.1523/JNEUROSCI.3355-13.2013

Kelly, S. P., & O'Connell, R. G. (2014). The neural processes underlying perceptual decision making in humans: Recent progress and future directions. J Physiol Paris. doi: 10.1016/j.jphysparis.2014.08.003

Kerns, J. G., Cohen, J. D., MacDonald, A. W., 3rd, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. Science, 303(5660), 1023-1026. doi: 10.1126/science.1089910

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J Neurophysiol, 97(6), 4296-4309. doi: 10.1152/jn.00024.2007

Kim, J. N., & Shadlen, M. N. (1999). Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat Neurosci, 2(2), 176-185. doi: 10.1038/5739

Kira, S., Yang, T., & Shadlen, M. N. (2015). A Neural Implementation of Wald's Sequential Probability Ratio Test. Neuron, 85(4), 861-873. doi: 10.1016/j.neuron.2015.01.007

Kolling, N., Behrens, T. E., Mars, R. B., & Rushworth, M. F. (2012). Neural mechanisms of foraging. Science, 336(6077), 95-98. doi: 10.1126/science.1216930

Kourtzi, Z., & Connor, C. E. (2011). Neural representations for object perception: structure, category, and adaptive coding. Annu Rev Neurosci, 34, 45-67. doi: 10.1146/annurev-neuro-060909-153218

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. J Neurosci, 20(9), 3310-3318.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. Science, 293(5534), 1506-1509. doi: 10.1126/science.1061133

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci, 2, 4. doi: 10.3389/neuro.06.004.2008

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., . . . Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron, 60(6), 1126-1141. doi: 10.1016/j.neuron.2008.10.043

Kunde, W. (2003). Sequential modulations of stimulus-response correspondence effects depend on awareness of response conflict. Psychon Bull Rev, 10(1), 198-205.

Lamb, T. D. (1980). Spontaneous quantal events induced in toad rods by pigment bleaching. Nature, 287(5780), 349-351.

Levi, D. M. (2008). Crowding--an essential bottleneck for object recognition: a mini-review. Vision Res, 48(5), 635-654. doi: 10.1016/j.visres.2007.12.009

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. Proc Natl Acad Sci U S A, 99(14), 9596-9601. doi: 10.1073/pnas.092277599

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. J Neurophysiol, 69(6), 1918-1929.

Liston, C., Matalon, S., Hare, T. A., Davidson, M. C., & Casey, B. J. (2006). Anterior cingulate and posterior parietal cortices are sensitive to dissociable forms of conflict in a task-switching paradigm. Neuron, 50(4), 643-653. doi: 10.1016/j.neuron.2006.04.015

Livne, T., & Sagi, D. (2007). Configuration influence on crowding. J Vis, 7(2), 4 1-12. doi: 10.1167/7.2.4

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. Curr Biol, 5(5), 552-563.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. Annu Rev Neurosci, 19, 577-621. doi: 10.1146/annurev.ne.19.030196.003045

Lupyan, G. (2013). The difficulties of executing simple algorithms: Why brains make mistakes computers don't. Cognition, 129(3), 615-636. doi: Doi 10.1016/J.Cognition.2013.08.015

Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. Annu Rev Neurosci, 37, 205-220. doi: 10.1146/annurev-neuro-071013-014017

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the Brain's Algorithm for Categorization from Its Neural Implementation. Current Biology, 23(20), 2023-2027. doi: Doi 10.1016/J.Cub.2013.08.035

Marchant, A. P., & de Fockert, J. W. (2009). Priming by the mean representation of a set. Q J Exp Psychol (Hove), 62(10), 1889-1895. doi: 10.1080/17470210902871045

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. Nature, 379(6566), 649-652. doi: 10.1038/379649a0

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. Nature Reviews Neuroscience, 4(4), 310-322. doi: Doi 10.1038/Nrn1076

McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. Nat Neurosci, 16(4), 493-498. doi: 10.1038/nn.3347

Medin, D. L., & Schaffer, M. M. (1978). Context Theory of Classification Learning. Psychol Rev, 85(3), 207-238. doi: Doi 10.1037//0033-295x.85.3.207

Meister, M. L., Hennig, J. A., & Huk, A. C. (2013). Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. J Neurosci, 33(6), 2254-2267. doi: 10.1523/JNEUROSCI.2984-12.2013

Messinger, A., Squire, L. R., Zola, S. M., & Albright, T. D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. Proc Natl Acad Sci U S A, 98(21), 12239-12244. doi: 10.1073/pnas.211431098

Milham, M. P., & Banich, M. T. (2005). Anterior cingulate cortex: an fMRI analysis of conflict specificity and functional differentiation. Hum Brain Mapp, 25(3), 328-335. doi: 10.1002/hbm.20110

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. Science, 263(5146), 520-522.

Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. Science, 254(5036), 1377-1379.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav Brain Res, 6(1), 57-77.

Mountcastle, V. B., LaMotte, R. H., & Carli, G. (1972). Detection thresholds for stimuli in humans and monkeys: comparison with threshold events in mechanoreceptive afferent nerve fibers innervating the monkey hand. J Neurophysiol, 35(1), 122-136.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. J Physiol, 283, 53-77.

Murphy, G. L. (2002). The big book of concepts. Cambridge, Mass. ; London: MIT Press.

Myczek, K., & Simons, D. J. (2008). Better than average: alternatives to statistical summary representations for rapid judgments of average size. Percept Psychophys, 70(5), 772-788.

Narayanan, N. S., Cavanagh, J. F., Frank, M. J., & Laubach, M. (2013). Common medial frontal mechanisms of adaptive control in humans and rodents. Nat Neurosci, 16(12), 1888-1895. doi: 10.1038/nn.3549

Navon, D. (1977). Forest before Trees - Precedence of Global Features in Visual-Perception. Cognitive Psychology, 9(3), 353-383. doi: Doi 10.1016/0010-0285(77)90012-3

Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). J Neurosci, 8(6), 2201-2211.

Nienborg, H., & Cumming, B. G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. Nature, 459(7243), 89-92. doi: 10.1038/nature07821

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. Journal of Experimental Psychology-General, 115(1), 39-57. doi: Doi 10.1037/0096-3445.115.1.39

Notebaert, W., & Verguts, T. (2006). Stimulus conflict predicts conflict adaptation in a numerical flanker task. Psychon Bull Rev, 13(6), 1078-1084.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nat Neurosci, 15(12), 1729-1735. doi: 10.1038/nn.3248

O'Neill, M., & Schultz, W. (2010). Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. Neuron, 68(4), 789-800. doi: 10.1016/j.neuron.2010.09.031

O'Neill, M., & Schultz, W. (2014). Economic risk coding by single neurons in the orbitofrontal cortex. J Physiol Paris. doi: 10.1016/j.jphysparis.2014.06.002

O'Reilly, J. X., Jbabdi, S., & Behrens, T. E. (2012). How can a Bayesian approach inform neuroscience? Eur J Neurosci, 35(7), 1169-1179. doi: 10.1111/j.1460-9568.2012.08010.x

O'Reilly, J. X., & Mars, R. B. (2011). Computational neuroimaging: localising Greek letters? Comment on Forstmann et al. Trends Cogn Sci, 15(10), 450. doi: 10.1016/j.tics.2011.07.012

O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. J Cogn Neurosci, 17(4), 580-590. doi: 10.1162/0898929053467550

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. Nat Neurosci, 4(12), 1244-1252. doi: 10.1038/nn767

Park, I. M., Meister, M. L., Huk, A. C., & Pillow, J. W. (2014). Encoding and decoding in parietal cortex during sensorimotor decision-making. Nat Neurosci, 17(10), 1395-1403. doi: 10.1038/nn.3800

Parker, A., & Hawken, M. (1985). Capabilities of monkey cortical cells in spatial-resolution tasks. J Opt Soc Am A, 2(7), 1101-1114.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. Annu Rev Neurosci, 21, 227-277. doi: 10.1146/annurev.neuro.21.1.227

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. Nat Neurosci, 4(7), 739-744. doi: 10.1038/89532

Philiastides, M. G., Biele, G., & Heekeren, H. R. (2010). A mechanistic account of value computation in the human brain. Proc Natl Acad Sci U S A, 107(20), 9430-9435. doi: 10.1073/pnas.1001732107

Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. Nature, 400(6741), 233-238. doi: 10.1038/22268

Posner, M. I., & Keele, S. W. (1968). On Genesis of Abstract Ideas. Journal of Experimental Psychology, 77(3P1), 353-&. doi: Doi 10.1037/H0025953

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. Nat Neurosci, 16(9), 1170-1178. doi: 10.1038/nn.3495

Rao, R. P. (2004). Bayesian computation in recurrent neural circuits. Neural Computation, 16(1), 1-38.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci, 2(1), 79-87. doi: 10.1038/4580

Ratcliff, R. (1978). Theory of Memory Retrieval. Psychological Review, 85(2), 59-108. doi: Doi 10.1037//0033-295x.85.2.59

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. Neural Computation, 20(4), 873-922. doi: Doi 10.1162/Neco.2008.12-06-420

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc Natl Acad Sci U S A, 106(16), 6539-6544. doi: 10.1073/pnas.0812589106

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. Psychol Rev, 106(2), 261-300.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. Curr Opin Neurobiol, 8(4), 516-521.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A, 97(22), 11800-11806. doi: 10.1073/pnas.97.22.11800

Reddy, L., & Kanwisher, N. (2006). Coding of visual objects in the ventral stream. Curr Opin Neurobiol, 16(4), 408-414. doi: 10.1016/j.conb.2006.06.004

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. Neuron, 61(2), 168-185. doi: 10.1016/j.neuron.2009.01.002

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic Distance and Verification of Semantic Relations. Journal of Verbal Learning and Verbal Behavior, 12(1), 1-20. doi: Doi 10.1016/S0022-5371(73)80056-8

Rishel, C. A., Huang, G., & Freedman, D. J. (2013). Independent category and spatial encoding in parietal cortex. Neuron, 77(5), 969-979. doi: 10.1016/j.neuron.2013.01.007

Rogers, T. T., & McClelland, J. L. (2008). Precis of Semantic Cognition: A Parallel Distributed Processing Approach. Behavioral and Brain Sciences, 31(6), 689-+. doi: Doi 10.1017/S0140525x0800589x

Romo, R., & Salinas, E. (2003). Flutter discrimination: neural codes, perception, memory and decision making. Nat Rev Neurosci, 4(3), 203-218. doi: 10.1038/nrn1058

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. Nat Neurosci, 5(7), 629-630. doi: 10.1038/nn866

Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., & Rushworth, M. F. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. J Neurosci, 28(51), 13775-13785. doi: 10.1523/JNEUROSCI.3541-08.2008

Rushworth, M. F., & Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci, 11(4), 389-397. doi: 10.1038/nn2066

Rushworth, M. F., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. Neuron, 70(6), 1054-1069. doi: 10.1016/j.neuron.2011.05.014

Saiki, J., & Holcombe, A. O. (2012). Blindness to a simultaneous change of all elements in a scene, unless there is a change in summary statistics. Journal of Vision, 12(3). doi: Artn 2 Doi 10.1167/12.3.2

Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. Nature, 354(6349), 152-155. doi: 10.1038/354152a0

Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. Nature, 346(6280), 174-177. doi: 10.1038/346174a0

Salzman, C. D., Murasugi, C. M., Britten, K. H., & Newsome, W. T. (1992). Microstimulation in visual area MT: effects on direction discrimination performance. J Neurosci, 12(6), 2331-2355.

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. Neuron, 20(2), 185-195.

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. J Cogn Neurosci, 26(8), 1736-1747. doi: 10.1162/jocn_a_00578

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. Curr Biol, 22(17), 1622-1627. doi: 10.1016/j.cub.2012.06.056

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. Nat Neurosci, 4(8), 819-825. doi: 10.1038/90526

Scolari, M., & Serences, J. T. (2009). Adaptive allocation of attentional gain. J Neurosci, 29(38), 11933-11942. doi: 10.1523/JNEUROSCI.5642-08.2009

Serences, J. T., & Saproo, S. (2012). Computational advances towards linking BOLD and behavior. Neuropsychologia, 50(4), 435-446. doi: 10.1016/j.neuropsychologia.2011.07.013

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol, 86(4), 1916-1936.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. Science, 237(4820), 1317-1323.

Sheth, S. A., Mian, M. K., Patel, S. R., Asaad, W. F., Williams, Z. M., Dougherty, D. D., . . . Eskandar, E. N. (2012). Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. Nature, 488(7410), 218-221. doi: 10.1038/nature11239

Siebert, W. M. (1965). Some implications of the stochastic behavior of primary auditory neurons. Kybernetik, 2(5), 206-215.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. Annu Rev Neurosci, 24, 1193-1216. doi: 10.1146/annurev.neuro.24.1.1193

Simons, D. J. (1996). In sight, out of mind: When object representations fail. Psychological Science, 7(5), 301-305. doi: Doi 10.1111/J.1467-9280.1996.Tb00378.X

Simons, D. J., & Levin, D. T. (1997). Change blindness. Trends Cogn Sci, 1(7), 261-267. doi: 10.1016/S1364-6613(97)01080-2

Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W., & Meister, M. (1997). Adaptation of retinal processing to image contrast and spatial scale. Nature, 386(6620), 69-73. doi: 10.1038/386069a0

Smith, E. E., & Medin, D. L. (1981). Categories and concepts. Cambridge, Mass ; London: Harvard University Press.

Soutschek, A., Taylor, P. C., Muller, H. J., & Schubert, T. (2013). Dissociable networks control conflict during perception and response selection: a transcranial magnetic stimulation Study. J Neurosci, 33(13), 5647-5654. doi: 10.1523/JNEUROSCI.4768-12.2013

Suchow, J. W., & Alvarez, G. A. (2011). Motion Silences Awareness of Visual Change. Current Biology, 21(2), 140-143. doi: Doi 10.1016/J.Cub.2010.12.019

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci, 11(9), 1004-1006. doi: 10.1038/nn.2163

Summerfield, C., & Tsetsos, K. (2012). Building Bridges between Perceptual and Economic Decision-Making: Neural and Computational Mechanisms. Front Neurosci, 6, 70. doi: 10.3389/fnins.2012.00070

Swets, J., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. Psychol Rev, 68, 301-340.

Talbot, W. H., Darian-Smith, I., Kornhuber, H. H., & Mountcastle, V. B. (1968). The sense of flutter-vibration: comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand. J Neurophysiol, 31(2), 301-334.

Tanaka, K. (1996). Inferotemporal cortex and object vision. Annu Rev Neurosci, 19, 109-139. doi: 10.1146/annurev.ne.19.030196.000545

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. Psychol Rev, 61(6), 401-409.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. Behav Brain Sci, 24(4), 629-640; discussion 652-791.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. Science, 331(6022), 1279-1285. doi: 10.1126/science.1192788

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. Nature, 381(6582), 520-522. doi: 10.1038/381520a0

Tipper, S. P. (1985). The negative priming effect: inhibitory priming by ignored objects. Q J Exp Psychol A, 37(4), 571-590.

Tosoni, A., Galati, G., Romani, G. L., & Corbetta, M. (2008). Sensory-motor mechanisms in human parietal cortex underlie arbitrary visual decisions. Nat Neurosci, 11(12), 1446-1453. doi: 10.1038/nn.2221

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. Psychol Rev, 95(1), 15-48.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. Cogn Psychol, 12(1), 97-136.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. Science, 247(4940), 301-306.

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: transfer across space and time. J Exp Psychol Hum Percept Perform, 35(1), 195-202. doi: 10.1037/0096-1523.35.1.195

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. J Cogn Neurosci, 21(10), 1934-1945. doi: 10.1162/jocn.2009.21131

Ullsperger, M., Bylsma, L. M., & Botvinick, M. M. (2005). The conflict adaptation effect: it's not just priming. Cogn Affect Behav Neurosci, 5(4), 467-472.

Ungerleider, L. G., Mishkin, M. (1982). Two Cortical Visual Systems. In G. M. M. R. Ingle DJ (Ed.), Analysis of Visual Behavior (pp. 549-586). Boston: MIT Press.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. Psychological Review, 108(3), 550-592. doi: Doi 10.1037//0033-295x.108.3.550

Vallbo AB, J. R. (1976). Skin mechanoreceptors in the human hand: neural and psychophysical thresholds. In Y. Zotterman (Ed.), Sensory Functions of the Skin in Primates (pp. 185-199). Oxford: Pergamon.

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. Neuroimage, 14(6), 1302-1308. doi: 10.1006/nimg.2001.0923

Vanegas, M. I., Blangero, A., & Kelly, S. P. (2013). Exploiting individual primary visual cortex geometry to boost steady state visual evoked potentials. J Neural Eng, 10(3), 036003. doi: 10.1088/1741-2560/10/3/036003

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. Ergonomics, 13(1), 37-58. doi: 10.1080/00140137008931117

Vogels, R., Sary, G., & Orban, G. A. (1995). How task-related are the responses of inferior temporal neurons? Vis Neurosci, 12(2), 207-214.

Wald, A., & Wolfowitz, J. (1948). Optimum Character of the Sequential Probability Ratio Test. Annals of Mathematical Statistics, 19(3), 326-339. doi: Doi 10.1214/Aoms/1177730197

Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron, 36(5), 955-968. doi: Doi 10.1016/S0896-6273(02)01092-9

Wark, B., Lundstrom, B. N., & Fairhall, A. (2007). Sensory adaptation. Curr Opin Neurobiol, 17(4), 423-429. doi: 10.1016/j.conb.2007.07.001

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. Vision Res, 32(5), 931-941.

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. Curr Opin Neurobiol, 8(2), 227-233.

Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. Vision Res, 24(1), 55-62.

Wunderlich, K., Rangel, A., & O'Doherty, J. P. (2009). Neural computations underlying action-based decision making in the human brain. Proc Natl Acad Sci U S A, 106(40), 17199-17204. doi: 10.1073/pnas.0901077106

Wyart, V., de Gardelle, V., Scholl, J., & Summerfield, C. (2012). Rhythmic fluctuations in evidence accumulation during decision making in the human brain. Neuron, 76(4), 847-858. doi: 10.1016/j.neuron.2012.09.015

Yang, T., & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. Nature, 447(7148), 1075-1080. doi: 10.1038/nature05852

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol Rev, 111(4), 931-959. doi: 10.1037/0033-295X.111.4.939

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. Neuron, 46(4), 681-692. doi: 10.1016/j.neuron.2005.04.026