Characterising feedback to mid-level visual cortex during perceptual decision-making

Dissertation

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät

und

der Medizinischen Fakultät

der Eberhard-Karls-Universität Tübingen

vorgelegt

von

Katrina Rose Quinn

aus Enfield, Großbritannien

2023

Tag der mündlichen Prüfung: 5.12.2023

Dekan der Math.-Nat. Fakultät: Prof. Dr. Thilo Stehle
Dekan der Medizinischen Fakultät: Prof. Dr. Bernd Pichler

1. Berichterstatter: Prof. Dr. Cornelius Schwarz
2. Berichterstatter: Dr. Hendrikje Nienborg

Prüfungskommission: Prof. Dr. Cornelius Schwarz
Dr. Hendrikje Nienborg
Prof. Dr. Ziad Hafed
Prof. Dr. Markus Siegel

Erklärung / Declaration:

Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:

„Characterising feedback to mid-level visual cortex during perceptual decision-making“ selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist be-kannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

I hereby declare that I have produced the work entitled “Characterising feedback to mid-level visual cortex during perceptual decision-making”, submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen, ...

Datum / Date Unterschrift / Signature

Statement of contributions

The work described in this thesis involved contributions from the doctoral candidate, Katrina R. Quinn, as well as other members of the Nienborg lab.

The work in chapter 5 is titled ‘A causal role for macaque area V2 in coarse disparity discrimination’. Katrina R. Quinn performed all of the analyses, generated the figures, and wrote the manuscript to be submitted. Hendrikje Nienborg and Bruce G. Cumming designed the research, and Hendrikje Nienborg collected the data.

The work described in chapter 6 titled ‘Decision-related feedback is spatially selective’ was published prior to the thesis. Therefore chapter 6 is at least partially identical with a manuscript of the publication Quinn et al. (2021). Katrina R. Quinn was involved in the training of two animals, collected the data, contributed to data analysis, generated the figures, wrote the first draft of the methods and supplementary material (which have here been integrated into the chapter contents), and co-wrote the remainder of the manuscript. Lenka Seillier trained one of the animals and contributed to data collection. Daniel Butts contributed to data analysis. Hendrikje Nienborg designed the research, contributed to data collection, performed data analysis and co-wrote the manuscript. The data described in this project is also used in a further manuscript currently under review (Talluri et al., 2022), entailing a co-authorship for Katrina R. Quinn who collected the data.

Arrakis teaches the attitude of the knife – chopping off what's incomplete and saying: 'Now, it's complete because it's ended here'

- Frank Herbert, Dune

Contents

1. Acknowledgements

Completing these projects was a huge undertaking, one that I couldn't have managed without the support of so many of my colleagues and friends.

I'd first like to thank my supervisor Hendrikje Nienborg. She trusted me with an important and challenging project, and guided me each step of the way. Throughout my time in the lab she encouraged me to seize every opportunity and not to shy away from the difficult elements of each task at hand. The lessons she taught me, particularly on presenting my work in the clearest possible way, were invaluable and continue to influence me today. It is no coincidence that I worked harder with her than I ever have before. For the encouragement and guidance, but also the challenges - thank you.

Working in a monkey lab is by no means a one-person job, and I have so much to be thankful for from my colleagues. Lenka Seillier, my lab mate from beginning to end, put enormous amounts of her time and energy into not only training the 2nd monkey but got involved in the recordings too – without her, this project could not have been finished. For that, and her continued companionship from monkey beginnings to heavily pregnant end, I am tremendously grateful. Paria Pourriahi laid the essential groundwork for our future recordings, all with a precision and grace that I desperately tried to emulate after she had left. She gifted me enormous technical knowledge, but most importantly taught me the monkey-whispering abilities that made our recordings possible. Katsuhisa Kawaguchi was always on hand to provide support, and never failed to cheer me with his humour and calm, sympathetic words. The work that Stephane Clery put into training the animals and showing me the ropes was essential to the project, but it was the friendship that he and Corinna Lorenz gave me that I will never forget. They brought so much joy, laughter and silliness to every day that we worked together, and continue to do so to this day. I'd also like to thank them and everyone else who has come through the lab, including those in the US, for their support, scientific discussions, and for listening to countless versions of my presentations. Lastly, a big thank you to Peter Dicke, for his technical assistance and willingness to answer any questions I had, no matter how small.

I couldn't go without mentioning the two animals that made all of this possible – our monkeys Kiwi and Mango. I didn't take the decision to work with these amazing

animals lightly, and the significance of their contribution to this work is not lost on me. I owe them everything. As for the fantastic team that took care of them, especially Manfred and Antje - it is their diligence and genuine affection for the animals that made working in the facilities such a positive experience.

I'd also like to thank all of those outside of work who supported me through the PhD. To all my friends, especially my Stiefelhof family in Tübingen, who were patient, kind and always happy to lend an ear. A huge and never-ending thank you to my parents: to Dad and Sarah, for the supportive chats, and much-needed distractions whenever I came to visit. To Mum and Simon - without their emotional and economic support I wouldn't have made it to Germany and I certainly wouldn't have the education that I have now. I'm sure mum regrets the former at least a little bit, but I am forever grateful.

And finally, to Florian, my partner in every sense of the word. Without you there is no way I would have made it. You have worn every hat imaginable in your quest to take care of and support me and I am in total awe of your capacity to encourage and inspire. Thankfully I have the rest of our lives together to pay you back.

2. Abstract

A long-standing question in neuroscience is how the activity of visual neurons supports perception. Historically examined from a purely feedforward perspective, this approach documented neuronal selectivity for specific perceptual features, sensitivity akin to an animal's perceptual sensitivity and demonstrated causal effects of sensory neurons on an animal's decision. Indeed, even the variable activity of single sensory neurons was found to be correlated with the decision an animal would make, often referred to as 'choice probability'. This decision-related activity was long interpreted as reflecting the causal effect of feedforward noise on the decision process, but increasing evidence has pointed to a feedback origin of these correlations with behaviour. However the role of that such feedback remains unclear. The work in this thesis sought to investigate the nature of this feedback in order to help explain what it's potential role in perceptual-decision making may be, as well as to further clarify long-held beliefs on the origin of decision-related activity. To do so, we focussed on the mechanisms underlying disparity perception in disparity-selective mid-level visual areas. First, we tested whether neurons in area V2 were causally involved in a disparity discrimination task. By electrically stimulating disparity-selective V2 neurons, we demonstrated a bias in the animals' decisions in line with the preference of the stimulated neurons, suggesting a causal role for these neurons in disparity perception. We then proceeded to better characterise the feedback that gives rise to decision-related activity in these neurons, as well as another group of disparity-selective neurons in V3/V3a. Since feedback has often been assumed to selectively target visual neurons based on their relevance for the task or stimulus demands, we aimed to test the extent of this selectivity. To do so, we employed a novel task combining disparity discrimination with a spatial attention component, wherein animals had to ignore one stimulus whilst discriminating the other. Critically, this led to distinct predictions for decision-related activity depending on how selective the feedback would be. We found that decision-related activity could be observed for neurons representing an ignored task-irrelevant stimulus, incompatible with accounts of feedback which exclusively target task-relevant neurons. Our findings suggest that decision-related activity arises predominantly as a result of feedback targeting neurons selective for disparity, regardless of whether they contribute to the task. Importantly they imply a biological constraint to the selectivity of feedback, and demand a revision of current theoretical accounts of feedback in perceptual decision-

making. The work presented here thus not only contributes to our understanding of disparity perception, but has critical implications for how feedback modulates the responses of visual neurons and ultimately shapes perception.

3. General Introduction

3.1. Selective visual neurons

For decades, visual neuroscientists have sought to explain how the brain is able to transform patterns of light falling on the retina into a meaningful percept of our world. Our understanding of the physiological basis of this process was fundamentally shaped by the pioneering work of Hubel and Wiesel in the 1950s and 60s (1959, 1962; see review from Wurtz, 2009). In the first of these experiments, electrodes were inserted into the striate cortex (V1) of awake behaving cats, and the responses of single neurons measured in action potentials or 'spikes' – popularly referred to as the currency of these cells. Using this setup, Hubel (1959) observed that simply waving a hand in front of the cat could illicit a flurry of activity in visual cortical neurons, moreso than diffuse illumination. To Hubel's surprise, through using more controlled light spots it became clear that the activity of these cells could be highly selective for the direction in which stimuli moved through their 'receptive field' (that is the confined area of the visual field within which light stimuli could modulate firing rate), firing more for one direction than another. Hubel & Wiesel (1959) used light slits to systematically measure this direction selectivity, as well as robust orientation selectivity, in V1 of the anaethestised cat, and later in the monkey (1962).

These early experiments paved the way for an extensive mapping of primate visual cortical selectivity for different visual features on an area-by-area basis. In addition to the direction and orientation selectivity in V1, selectivity for direction was also observed in MT (e.g. Albright, 1984; Maunsell & van Essen, 1983; Zeki, 1974) and MST (Desimone & Ungerleider, 1986), for orientation in V2 (Burkhalter & Van Essen, 1986; Hubel & Livingstone, 1987), V3 and V3a (Zeki, 1978), colour in V2 (Burkhalter & Van Essen, 1986; Hubel & Livingstone, 1987) and V4 (Zeki, 1978), and binocular disparity in V1 (G. F. Poggio & Fischer, 1977), V2 (Hubel & Livingstone, 1987), V3, V3a (Anzai et al., 2011) and MT (J. H. Maunsell & Van Essen, 1983), to name some examples. Further, in a number of these areas, selectivity for specific features appeared to be organised topographically, either in columns or clustered. Well-known examples include the orientation columns (Hubel & Wiesel, 1968) and colour blobs of V1 (Livingstone & Hubel, 1984), but additional examples include direction selective

columns in MT (Albright et al., 1984), and disparity clusters in area V2 (Chen et al., 2008). This spatial organisation of neurons according to their feature selectivity suggested that these areas may be specialised for the processing of the given feature.

3.2. The role of visual neurons in perception

Having established neuronal selectivity for a range of visual features, the next step was to test whether these neurons contributed to the perception of said features. This drove a wave of experiments which sought to directly compare the selectivity of visual neurons with the behavioural performance of animals, by using psychophysical tasks tailored to the selectivity of the neurons under study. To begin with this was indirect – comparing behavioural observations in humans with neural recordings in animals (Newsome et al., 1986; Tolhurst et al., 1983). However, it was through the work of Newsome and colleagues that monkeys were trained to perform the psychophysical tasks themselves, facilitating a direct comparison of visual neuronal activity and behaviour that had not previously been possible (Britten et al., 1992; Britten et al., 1996; Celebrini & Newsome, 1994; Newsome et al., 1989; see review from Parker & Newsome, 1998). In these experiments (see Fig. 1), monkeys were trained to perform a motion discrimination task, in which the animal had to determine whether they perceived one or the other direction of motion in a noisy random-dot stimulus (e.g. up vs. down) and report their choice with a motor response. These stimuli allowed the experimenters to carefully titrate the degree of signal, in this case the net motion, to the lowest levels at which the observer could reliably discriminate the direction presented. This could be achieved by manipulating the proportion of dots moving in an experimentally specified direction. For each signal level, and across multiple repetitions, one could then measure the proportion of choices in favour of a given direction. For the highest included signal levels, for each direction, an animal's choices reached near-perfect performance, but as the signal decreased, the choices became more variable. Importantly, a portion of the presented stimuli had no net motion, or zero signal, appearing to the observer as "twinkling visual noise" (Newsome & Pare, 1988). Even on these zero signal trials the animals were still required to report the motion direction they perceived, and were rewarded randomly for their choice. This meant that the choices on these trials could be thought of as largely independent of the sensory stimulus, the importance of which will become apparent in the next section (see 3.4 Decision-related Activity).

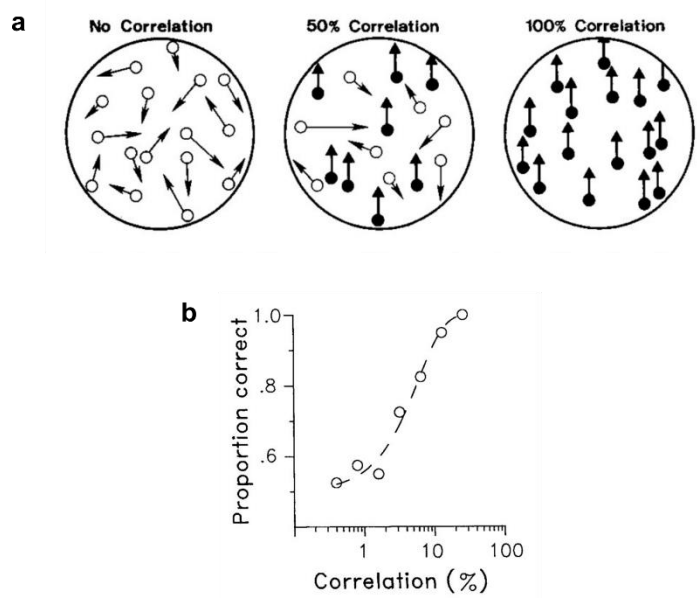

Fig. 1 | Example motion discrimination task. a) Stimulus schematic. In Britten et al., (1992) motion coherence was determined by the percentage ("correlation") of dots moving in the same direction. At no correlation (left) all dots were randomly replotted and appeared as noise. The strength of the motion signal increased with increasing correlation. **b)** Example psychophysical performance. As motion signal (correlation, %) increased so too did the proportion of correct responses. Data fit with a psychometric function (sigmoid function). Figure adapted from "The Analysis of Visual Motion: A Comparison of Neuronal and Psychophysical Performance" by K.H. Britten, M.N. Shadlen, W.T. Newsome and J.A. Movshon, 1992, *Journal of Neuroscience*, **12**, p.4746. Copyright 1992 by the Society for Neuroscience.

In order to make any comparison between these behavioural measurements and the neural recordings, both the behaviour and neural data needed to be quantified in a comparable way. For behaviour, the proportion of choices in favour of one direction was plotted as a function of the signal values for both directions and fit with a sigmoid curve, yielding what is referred to as the psychometric function. The psychometric function describes the relationship between the animal's perceptual choices and the strength of the stimulus signal (Treutwein & Strasburger, 1999; Wichmann & Jäkel, 2018), and it's fit parameters allow for the quantification of specific aspects of the animal's discriminatory capabilities. For example, the threshold (when defined as the signal value at 50% correct choices) could tell us the signal level at which the animal is equally likely to report one or the other direction (point of subjective equality, PSE). In addition, the slope around this threshold gives us a measure of how sensitive their

behaviour is to small differences in signal levels. Importantly, the psychometric function provides a standardised measure through which an animal's psychophysical behaviour can be compared throughout task-learning and across task conditions or animals.

To compare this psychometric function with the neural data, a similar standardised measure of neural sensitivity was devised called the neurometric function (Newsome et al., 1989). To do so, an approach from signal detection theory was utilised (Green & Swets, 1966), known as the receiver operating characteristic (ROC) curve (Fig. 2). In short, say we plot the distribution of firing rates before a preferred and a non-preferred or 'null' stimulus (Fig. 2a). For a given firing rate threshold we can then observe how often a particular firing rate above/to the right of the threshold occurred in response to a preferred (true positive) or null stimulus (false positive). Conversely, for firing rates under/to the left of the threshold we can measure true negatives (null stimulus presented) and false negatives (preferred stimulus presented). Using these four measures, one can calculate the true positive and false positive rates (TPR and FPR respectively). The ROC curve is then plotted as the FPR against TPR for a range of thresholds/criterion (Fig. 2b), and the resulting area under the curve (aROC) can be used as a measure of the discriminability of the signals of interest. In this particular case, how well do the neuron's responses discriminate between preferred and null stimuli. The neurometric curve itself consisted of plotting the aROC value as a function of signal level (Fig. 2c), which can be fit with a sigmoid curve.

Using a motion discrimination task and simultaneous recordings in single MT and MST neurons, Newsome and colleagues found the psychometric and neurometric functions to be stikingly similar in both the average threshold and slope (Britten et al., 1992; Celebrini & Newsome, 1994; Newsome et al., 1989). This suggested that the sensitivity of individual visual neurons was sufficient to account for an animal's perceptual behaviour, and implied that these neurons may play a key role in perception.

3.3. Establishing a causal role

The feature selectivity of visual neurons and the similarity between neuronal and behavioural sensitivity during a perceptual task provided substantial evidence that these neurons could be involved in visual perception. However, this did not provide evidence for a causal role of these neurons in perception. To do so, perceptual

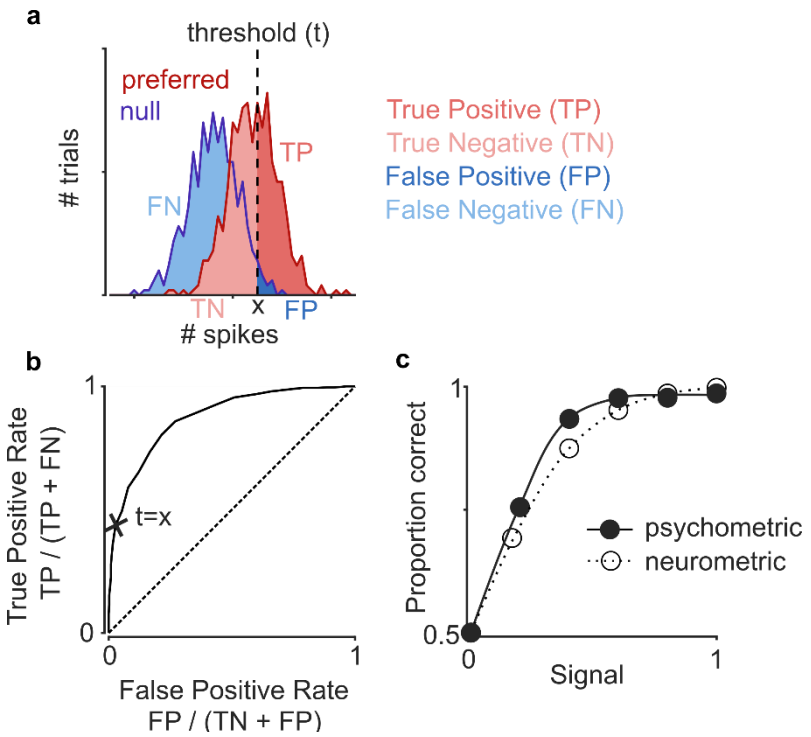

Fig 2. Illustration of ROC analysis. a) Hypothetical firing rate distributions for a single neuron to preferred (red) and null (blue) stimuli based on simulated data. Decision threshold (dashed black line) determines whether a given firing rate is taken to indicate a preferred (right of threshold) or null (left of threshold) stimulus. Firing rates can subsequently be grouped into one of four categories based on this threshold and the presented stimulus (TP, FP, TN, FN). **b)** Receiver operating characteristic (ROC) curve. For a range of threshold values, the false positive rate (FPR) is plotted against the true positive rate (TPR). Black cross shows point on curve corresponding to a threshold of x (shown in a). The area under the resulting curve (aROC) corresponds to how well the preferred and null firing rate distributions can be separated. **c)** Simulated psychophysical performance for a discrimination task (solid line for illustrated fit and solid circles for individual data points) as degree of signal (e.g. % correlated motion) by proportion correct choices. Neural performance for the same task based on hypothetical single neuron (dashed line for fit, open circles for data points). Here proportion correct corresponds to the aROC value i.e. how well the neuron discriminates between preferred and null stimuli at each signal level. Note the similarity between psychophysical and neural performance based on previous findings (see main text).

experiments needed to be combined with manipulations that could artificially perturb the activity of visual neurons, in order to measure the effect on behaviour.

One method to achieve this is by lesioning a particular region in order to observe whether this leads to a selective impairment of the animal's behaviour in a task requiring perception of a given feature (e.g. Newsome & Pare, 1988; Orban et al., 1995). Indeed, Newsome and Pare (1988) found that lesioning area MT led to a

detriment in behavioural performance in the motion discrimination task, but not in an orientation discrimination task.

An alternative approach is to leave the region of interest intact, and instead electrically stimulate a cluster of neurons with a shared preference for a particular feature (see review from Cohen & Newsome, 2004), known as microstimulation. The stimulation is thought to act as if adding to the signal in support of the neurons' preferred feature. Thus, if microstimulation leads to an increase in reports for the neurons' preferred feature, this would go some way towards establishing a causal role for these neurons. This is precisely what was observed by Salzman et al. (1992; 1990), who found that stimulation of MT neurons led to an increase in choices in favour of the direction preference of the stimulated neurons.

The results of both of these approaches came together to suggest a causal contribution of neurons in area MT, which marked a significant turning point in the understanding of their role in the perception of motion direction. Indeed perturbation techniques like microstimulation (but increasingly optogenetics, see Galvan et al., 2017) continue to be used as a tool to test the causal role of a particular group of neurons in perception (Cohen & Newsome, 2004; Wurtz, 2015), which we utilise in chapter 5 of the thesis.

3.4. Decision-related activity

Whilst the comparisons of neural and behavioural sensitivity had been made at the single neuron level, the above causal manipulations focussed on the effects of perturbing a large population of visual neurons simultaneously, establishing a causal role for these selective neuronal populations in a motion discrimination task. However, even at the single neuron level, an intriguing correlation could be observed between stimulus-independent activity fluctuations and the decision an animal would make in this task (Britten et al., 1996). Concretely, if a direction-selective neuron fired more on a given trial, there was a greater likelihood that the animal would make a choice towards that neuron's preferred direction. Importantly, this activity was stimulus-independent, in that it could be measured for repeated presentations of identical stimuli, including those without any net motion signal. This decision-related activity was quantified by Britten and colleagues as 'choice probability', given as the area under the ROC curve over the firing rate distributions before a preferred vs. null choice (same

analysis as illustrated in Fig. 2a,b but for choice rather than stimulus elicited firing rates). The choice probability metric yielded a value between 0 and 1, corresponding to the probability with which an observer could predict the choice an animal would make, given the preference of the neuron and knowledge of the firing rate distributions before each choice. Values deviating from 0.5 signified performance better than chance. Although the average choice probabilities found were small (0.55, Britten et al., 1996), they demonstrated that small fluctuations in neural activity unrelated to the stimulus could be systematically associated with the choice an animal would make.

These results have since been extended to a wide range of different visual features and areas (see review from Crapse & Basso, 2015) although here we will focus on this activity in purely visual neurons, observed for example for orientation in V1 (Nienborg & Cumming, 2014), disparity in V2 (Clery et al., 2017; Nienborg & Cumming, 2006), MT (Uka & DeAngelis, 2004) and IT (Uka, 2005), and structure-from-motion in MT (Dodd et al., 2001). Furthermore, the magnitude of these trial-by-trial fluctuations with choice often depends on how well-tuned the neuron is for the task (e.g. Britten et al., 1996; Uka & DeAngelis, 2004; Nienborg & Cumming, 2014), with higher choice probabilities for neurons tuned more strongly for the task-relevant feature.

At first glance, choice probabilities appeared to provide a metric through which to gauge the involvement of the measured neurons to the decision, the magnitude of which would then reflect the weight given to that neuron in the readout of choice from the sensory population. However, as we shall see in the remaining sections, this simple interpretation ignored that this correlational metric need not reflect causation, and that its magnitude, rather than reflecting readout weight, might reflect the relationship of the measured single neuron with a wider pool of sensory neurons.

3.5. The importance of correlated variability for decision-related activity

That choice probabilities of the observed magnitude could be measured in single neurons at first puzzled researchers (Zohary et al., 1994). If the measured neuron contributed as much as it's correlation suggested, then presumably only a very small pool of neurons could be responsible for the perceptual decision. Even using stimuli optimised for the neurons under study, how would the experimenter get lucky enough

to find one of the few neurons that contributed to the percept out of many hundreds with similar tuning?

The answer is correlated variability. It had already been observed that the activity of visual neurons fluctuated even in response to repeated presentations of identical visual stimuli (Tolhurst et al., 1983). When measuring these variable responses simultaneously for pairs of MT neurons during passive viewing of motion stimuli or the motion discrimination task, Zohary et al. (1994) found that the activity fluctuations were correlated between the neuronal pairs. These pairwise correlations are often referred to as correlated variability, or noise correlations. Correlated variability was shown to be higher between neurons with similar tuning (Zohary et al., 1994; Bair et al, 2001; Smith & Kohn, 2008) and to decrease with increasing neuronal distance (Smith & Kohn, 2008), being close to zero for neurons across hemispheres (Cohen & Maunsell, 2009). Whilst mean correlation values between 0.1 and 0.2 have typically been reported (e.g. Kohn & Smith, 2005; Bair et al., 2001; Zohary et al., 1994; see review from Cohen & Kohn, 2011), it will become clear that the exact magnitude of these correlations is not the relevant factor for understanding their relationship with decision-related activity.

The importance of correlated variability for decision-related activity was first made explicit in a model from Shadlen and colleagues (1996) which remains influential to this day. The authors sought to explain several of the findings described above documenting correlations between the activity of sensory neurons with behaviour, as well as with other sensory neurons, during a motion discrimination task. In the model, decisions were made by comparing the average activity of two pools of neurons supporting opposite choices (e.g. up vs. down motion), whose responses were based on those of real MT neurons (Britten et al., 1996). The decision was then made in favour of the choice corresponding to the pool with the greatest average activity. The success of the model was gauged on whether it was able to reproduce decision-related activity and psychophysical thresholds of the previously observed magnitudes, whilst manipulating variables of pool size, tuning composition, pooling noise and within-pool correlations. Using biologically plausible pool sizes and a broad range of neuronal selectivity, the authors found that the magnitude of decision-related activity depended on weak pairwise correlations between neurons within a pool, with decision-related activity increasing with increasing correlated variability. Practically, this means that a

single neuron's decision-related activity would not simply reflect it's own contribution to the decision, but also the contribution of all neuron's within the same pool (Cohen & Kohn, 2011; Haefner et al., 2013). Indeed, some neurons could even be correlated with choice simply by virtue of their being correlated with neurons that do contribute, even if they themselves were not given any weight in the decision. This not only explains the puzzle of why it is relatively easy to find single neuron's that exhibit decision-related activity as large as has been observed, but also drives home why this metric cannot be used to infer a causal relationship with behaviour.

Whilst Shadlen and colleagues' model highlighted the importance of within-pool correlations for decision-related activity, it did not consider the impact of between-pool correlations, which were held at 0. These alone would not impact decision-related activity in this model as any activity shared between pools would be subtracted out at the decision-stage. However, the difference between within- and between-pool correlations would have critical effects on the magnitude of decision-related activity, with higher values resulting from larger within-pool vs. between-pool correlations (Nienborg & Cumming, 2010). Thus it is the structure of correlated variability, rather than the magnitude, which has important implications for decision-related activity. Although the relationship between correlated variability and tuning similarity would seem to support this structure, we shall later see that the structure is far less fixed than initially believed.

It should also be noted that not all correlated variability has implications for decision-related activity, but rather a specific subtype – that is variability which appears to mimic a change in the stimulus (Kanitscheider et al., 2015; Moreno-Bote et al., 2014; Pitkow et al., 2015), so-called 'information-limiting' noise correlations.

3.6. Generalisation across features

Whilst many of the results discussed above focused on the role of MT neurons in motion perception, there is no reason to believe these findings are unique to this particular feature or area (as alluded to in several subsections). For these experiments, one could take any number of visual features and the corresponding selective neuronal clusters to facilitate our understanding of the role of visual neurons in perception.

3.7. Binocular disparity

In this thesis, we chose to examine the perception of binocular disparity in mid-level visual cortex. Binocular disparity refers to the small differences in the position of an object between the eyes, providing an essential cue for depth perception (Cumming & DeAngelis, 2001; DeAngelis, 2000). Selectivity for binocular disparity can be found as early as area V1 (Poggio & Fischer, 1977) where information from both eyes first converges upon the same single neurons (DeAngelis, 2000), as well as in a number of extra-striate regions including V2, V3, V3a and MT (Anzai et al., 2011; Hubel & Livingstone, 1987; Maunsell & Van Essen, 1983; Poggio & Fischer, 1977)

Similar to the motion discrimination tasks discussed, typical disparity discrimination tasks also utilise random-dot stimuli, but instead of moving, a portion of these dots are horizontally displaced in the image sent to one eye with respect to those sent to the other eye. This displacement, or disparity, creates an illusion of depth wherein a portion of the stimulus appears to either protrude or recede relative to the surround (often the plane of fixation), depending on the direction of displacement. The animals' task is then to report whether this displaced region is 'near' (protruding) or 'far' (receding) with respect to the surround (e.g. DeAngelis et al., 1998; Nienborg & Cumming, 2006; Nienborg & Cumming, 2009). During such a task, decision-related activity has been observed in V2 and MT (Nienborg & Cumming, 2006; Nienborg & Cumming, 2009; Uka & DeAngelis, 2004), but notably not in V1 (Nienborg & Cumming, 2006). This is believed to be due to a difference in the spatial organisation of disparity selectivity within these areas (Nienborg & Cumming, 2014) – whilst V1 exhibits no topographical representation for disparity, neurons in both V2 and MT appear to be clustered according to their disparity preference (Chen et al., 2008; DeAngelis & Newsome, 1999a). Given that V2 is the earliest cortical area to exhibit such topographical organisation for disparity suggests it may indeed play an important role in disparity perception. However no previous study has attempted to show a causal role for V2 neurons in disparity perception. In chapter 5 of this thesis we therefore sought to rectify this omission by using microstimulation to establish a causal role for V2 neurons in disparity perception during a disparity discrimination task. In chapter 6, also using a disparity discrimination task, our recordings were focussed on area V2, as well as area V3/V3a, which receives input from V2 (Felleman et al., 1997) and has

likewise shown substantial disparity selectivity (Adams & Zeki, 2001; Anzai et al., 2011; Hubel et al., 2015).

3.8. A feedforward approach

The contents of this thesis thusfar have described a strictly feedforward approach to understanding perception. Stimuli were designed specifically with the selectivity of the measured neurons in mind, e.g. random-dot stimuli with motion for motion-selective neurons, or with disparity for disparity-selective neurons, and used in tasks that sought to isolate feedforward processing. The resulting observations, including decision-related activity, could be accounted for by a simple pooling in which the flow of information goes in one direction – fed forward (Shadlen et al., 1996). Consequently, decision-related activity has long been interpreted as reflecting the causal effect of variable neural responses on the choice.

However as discussed in the section on correlated variability, the interpretation of decision-related activity is not clear cut. Firstly, the dependence on correlated variability means that decision-related activity likely arises at least in part due to correlations with other neurons, such that a single neuron correlated with the decision need not be causally involved in the decision. Secondly, a specific structure of this correlated variability appears to be essential to observing decision-related activity, i.e. higher within vs. across pool correlations. Thus the origin of this correlated variability becomes a critical question in our understanding of decision-related activity and its relationship to perception.

Such correlated variability between visual neurons is commonly believed to arise through shared inputs. Initially these were assumed to be exclusively feedforward inputs (Shadlen et al., 1996). This assumption was subsequently supported by a recent feedforward model of neural activity in V1, showing that shared noise from the sensory periphery was sufficient to account for information-limiting noise correlations in the context of an orientation discrimination task (Kanitscheider et al., 2015). Still, shared inputs need not arise from feedforward sources only, especially considering the extensive feedback connections within the visual system (Briggs, 2020). If correlated variability instead arose from shared feedback inputs, this would not necessitate fundamental changes to the structure of the decision-making models discussed above, but could potentially have critical implications for our interpretation of decision-related

activity. Thus, in order to better understand decision-related activity, it is critical to uncover the origin of shared variability (Cumming & Nienborg, 2016). As will become apparent in the following section, as well as in our own findings presented in chapter 6 of the thesis, there is intriguing evidence for a feedback origin of so-called noise correlations.

3.9. Feedback modulation of visual neurons

Whilst the feedforward approach to perception yielded significant insights into the potential mechanisms by which visual neurons give rise to perception, it ignored a fundamental property of the visual cortex - that it is highly recurrent, with large numbers of lateral and feedback connections (see review from Briggs, 2020). Thus, the question of how visual neurons shape perception generally needs to consider whether feedback may play some role in this process.

Indeed, there is ample evidence that feedback modulates the activity of sensory neurons (see reviews from Nienborg & Roelfsema, 2015; Petro et al., 2014; Roelfsema & de Lange, 2016). Here we focus on a particularly well-studied example which will be relevant for the remainder of the thesis – attentional modulation. It has been repeatedly observed that attending to a particular spatial location or stimulus feature, leads to a robust increase in the firing rates of visual neurons selective for the attended location or feature (Desimone & Duncan, 1995; Treue & Maunsell, 1996; see review from Reynolds & Chelazzi, 2004;). In a typical spatial attention task, the observer is required to attend to a stimulus at a particular location in the visual field, leading to an upregulation of the responses of those neurons whose receptive fields overlap with that location (e.g. Treue & Maunsell, 1996). Furthermore allocation of spatial attention has also been shown to lead to a reduction in correlated variability between pairs of visual neurons (Cohen & Maunsell, 2009; Mitchell et al., 2009). The effects of feature attention can be demonstrated in similar tasks (see review from Maunsell & Treue, 2006; Treue & Trujillo, 1999) but are perhaps best explained in the context of visual search, where observers are required to search for a particular target amongst an array of distractors, using a specific feature of that stimulus, for example its colour (Bichot et al., 2005). In such a task, when an animal attends to a specific feature, such as the colour red, the responses of neurons with a preference for the colour red are modulated, resulting in an increase in firing rate (Bichot et al., 2005). Interestingly,

although the effects of spatial attention are restricted to neurons whose receptive fields overlap with the attended spatial location, feature attention has been shown to act on neurons across the visual field which show a preference for the attended stimulus feature (Bichot et al., 2005; Treue & Trujillo, 1999).

3.10. Decision-related feedback

Even for the decision-making paradigms previously mentioned, which had been developed specifically to isolate the feedforward contribution of visual neurons to perception, there is increasing evidence that feedback modulates neural responses (Bondy et al., 2018; Cohen & Newsome, 2008; Nienborg & Cumming, 2009; Wimmer et al., 2015) and can at least in part account for the observed decision-related activity (Bondy et al., 2018; Nienborg & Cumming, 2009; Wimmer et al., 2015). One prominent example comes from a study from Nienborg and Cumming (2009), in which the authors compared the timecourse of the animal's perceptual behaviour in a disparity discrimination task with that of the decision-related activity in area V2. They found that although the animal gave progressively less weight to the stimulus evidence as the trial progressed, decision-related activity conversely increased and remained high towards the end of the trial. If decision-related activity arose solely as a consequence of the measured neurons' causal effect on the decision, it would be expected to decrease as the animal's use of the stimulus information decreased. Thus, this mismatch between the behavioural and neural timecourses suggested that decision-related activity was not simply a consequence of the feedforward influence of neurons on the decision, arguing for a feedback component to this activity. Similarly, a late component of decision-related activity was observed by Wimmer and colleagues (2015), and could be accounted for by feedback to sensory neurons as the animals made their decision.

There is likewise evidence that correlated variability can be influenced by feedback. As referenced above, attention to a particular stimulus leads to a reduction in noise correlations between neurons in area V4 whose receptive field overlaps with this stimulus (Cohen & Maunsell, 2009; Mitchell et al., 2009). More importantly, the structure of noise correlations seems to change with the task structure, suggesting that at least part of the correlated variability observed in visual neurons may have a feedback origin (Bondy et al., 2018; Cohen & Newsome, 2008). In an elegantly

designed experiment from Cohen and Newsome (2008), the authors measured from pairs of neurons in area MT whilst animals performed a motion discrimination task. On a trial-by-trial basis, the experimenters changed the axis along which motion needed to be discriminated, e.g. upwards from downwards motion on one trial, and leftwards from rightwards motion on another. They found that noise correlations between the same pair of neurons changed depending on whether these two neurons supported the same or different decisions, with higher values for neurons supporting the same decision. Similar findings have since been observed in V1 in the context of an orientation discrimination task, again with the magnitude of noise correlations depending on whether neurons supported the same or different decisions (Bondy et al., 2018).

These findings drive home the importance of the within vs. across pool structure discussed previously, but importantly offer an alternative explanation of their origin. That a component of this structure was not fixed, but instead changed in accordance with the task-relevant motion direction, indicates that it at least in part arises as a result of feedback. More specifically, simulations from Cohen and Newsome suggest that this structure could arise from a similar mechanism to that of feature attention, an idea which has been echoed in the literature (e.g. Bondy et al., 2018; Krug, 2004; Nienborg & Cumming, 2009). To illustrate this, consider the example of a motion discrimination task. On any given trial, an animal attends to the stimulus and must discern what motion direction they perceive. If the animal begins to attend to one of the possible task-relevant motion directions, for example upwards motion, this would, in line with what we know of feature attention, lead to an upregulation of the responses of those neurons with a preference for upwards motion. As these neurons would share this feedback input, their responses would in turn become more correlated, compared to neurons with different preferences not receiving this feedback. Further, as long as the attended feature systematically relates to the choice the animal makes, this would also give rise to decision-related activity. Indeed, Bondy and colleagues (2018) found that task-dependent changes in noise correlation structure could account for the decision-related activity they observed. These findings, and the potential mechanism, are thus also compatible with the feedback component of decision-related activity described above. These results call into question the long-dominant interpretation of decision-related activity as reflecting causal involvement of sensory noise on the decision

process (Crapse & Basso, 2015; Nienborg & Cumming, 2010). That these phenomena can arise through feedback furthermore highlights the importance of using perturbation techniques, such as microstimulation, to test for causal relationships between particular neurons and perception (as in chapter 5).

3.11. The role of feedback in perception

In spite of this growing evidence for feedback modulation of visual neurons during perceptual decision-making, its precise role is still unclear. A common suggestion is that feedback relays important information about the task or stimulus context, or expectations of the incoming sensory evidence, in order to help prioritise the processing of task-relevant stimuli or features (Harris & Mrsic-Flogel, 2013; Nienborg & Roelfsema, 2015; Petro et al., 2014; Roelfsema & de Lange, 2016). Theoretical models of perception have elaborated on this concept, proposing precise computational roles for feedback (Haefner et al., 2016; Haimerl et al., 2019; Rao & Ballard, 1999). The strength of these models is that they make specific predictions for commonly observed response modulations, such as decision-related activity and the structure of noise correlations, that can then be tested experimentally. For example, one recent compelling model from Haefner and colleagues (Haefner et al., 2016) linked the activity of sensory neurons to a historical concept of perception as inference (Helmholtz, 1867), in which the currently available sensory evidence is combined with previously acquired knowledge or expectations of the world. In their framework of probabilistic inference by neural sampling, sensory neurons represent an image, or stimulus, based on samples. This leads to the formation of a task-specific belief about the stimulus, which then informs the decision variable. Critically, this belief feeds back to the sensory neurons, creating a feedback- or confirmation loop between the sensory neurons and the decision variable that would account for the previously observed late component of decision-related activity (Nienborg & Cumming, 2009; Wimmer et al., 2015). The model further predicted that once the animal had learned the task, the feedback of this belief would result in noise correlations structured with respect to the task-relevant features in line with experimental observations (Bondy et al., 2018; Cohen & Newsome, 2008). The structure of noise correlations in this framework would thus arise if the belief selectively targeted those neurons tuned for the task-relevant features, fluctuating on each trial in accordance with the belief.

This idea of selective targeting of task-relevant neurons is not unique to the model from Haefner and colleagues, instead forming a critical part of other models of perceptual decision-making - for example, by selectively tagging the relevant neurons to facilitate downstream readout (Haimerl et al., 2019) - and perceptual learning (Roelfsema et al., 2010). In fact, one could argue that any concept of feedback as a conveyor of task-relevant context or expectations implies that this feedback selectively target task-relevant neurons. If feedback blanket modulated neurons across visual cortex, without regard for their selectivity or their role in the task, this would simply act to enhance responses overall. Importantly, such unselective feedback would not give rise to the feedback modulations we have described above, including for attention, noise correlations and decision-related activity. The proposed roles for feedback and the ability to account for these modulations therefore rely on some degree of selectivity, regardless of the specific computational role.

This then poses the question of just how selective feedback to visual neurons is. Current theoretical accounts (e.g. Haefner et al., 2016; Haimerl et al., 2019) seem to suggest feedback to be as selective as the task demands. Although there is anatomical evidence for some degree of selectivity (Briggs, 2020; Federer et al., 2021), given the number of possible tasks we face, the anatomical wiring that would be required for maximal task-selectivity would be biologically costly. Understanding potential limits of feedback selectivity could provide valuable constraints for theoretical accounts of the computational role that feedback plays in perception, and become an important touchstone from which to better understand how the activity of visual neurons informs perception. Therefore our main question in chapter 6 of the thesis was to probe the selectivity of feedback in perceptual decision-making.

4. Research Aims

The overall aim of this thesis was to better characterise the mechanisms through which visual neurons give rise to perception, building on the work discussed above. To do so, we focussed on disparity perception and the activity of disparity-selective neurons in mid-level visual areas.

In chapter 5, we first wanted to establish a causal role of neurons in area V2 to disparity perception. The focus was on V2 as this area is the first in the cortical hierarchy to show a topographical organisation for disparity, as well as decision-related activity in a disparity discrimination task. However, no previous studies had demonstrated a causal role for these neurons in the perception of disparity. We therefore sought to test this, by applying microstimulation to disparity selective V2 neurons whilst animals performed a disparity discrimination task. If neurons in V2 did contribute to the perception of disparity in this task, we would expect microstimulation to lead to an increased proportion of choices towards the preference of the stimulated neurons.

Equipped with the knowledge of whether V2 neurons are causally involved in disparity perception, in chapter 6 I could then go on to probe the selectivity of feedback to neurons in V2, as well as in disparity-selective neurons in V3/V3a, in a disparity discrimination task. We considered the evidence highlighted above: both empirical and theoretical work suggests that feedback selectively targets task-relevant neurons. In a feature discrimination task, like disparity discrimination, this would involve targeting those neurons selective for the discriminated feature. As discussed, such feature-selective feedback was able to account for changes in the structure of noise correlations, and the feedback component of decision-related activity. Further, this feature-selective feedback has repeatedly been suggested to employ the same mechanism as that of feature attention. As briefly mentioned, one known characteristic of feature attention is that it acts in a spatially global manner, targeting all neurons with a preference for the relevant feature regardless of their spatial location (Bichot et al., 2005; Treue & Trujillo, 1999). Whilst such global modulation would be beneficial in a task in which the location of the task-relevant stimulus is unknown, such as in visual search, it provides no obvious advantage in a discrimination task, which typically involves a single stimulus presented at a specified location. With this knowledge we

were able to design a task which would generate distinct predictions for decision-related activity, depending on how selective the feedback was. Specifically we employed a novel task combining disparity discrimination with an added spatial attention component. In addition, this task in combination with across-hemisphere recordings would also allow us to measure noise correlations within and across hemispheres. This provided an important control to confirm that our results were not driven by similarities in tuning unrelated to feedback, as well as supporting previous findings that show task-related changes in the structure of noise correlations (Bondy et al., 2018; Cohen & Newsome, 2008).

Together, the research questions posed in these chapters should contribute to the rich literature describing the role of visual neurons in perception. Establishing the causal role of V2 in disparity perception would not only help in building a more complete understanding of how particular features are processed through the visual system, but would also provide essential knowledge to facilitate the selective targeting of causally involved neurons. Indeed, we capitalised on the latter in our investigation of the selectivity of decision-related feedback. Demonstrating that decision-related feedback is not spatially-selective, and instead acts in a global manner, would have at least two important implications. First, it would help provide evidence for the previous suggestion that decision-related activity and feature-selective attention share a common mechanism, bringing together two largely independent fields of research into a common framework which could facilitate better understanding of both. Secondly, it could provide a critical insight into how feedback may generally be deployed across the brain. Specifically, rather than being completely tailored for the current task demands, feedback may be limited in it's selectivity and capitalise on a smaller set of base processes or channels. By doing so, the brain could minimise the biological costs of extensive wiring as well as facilitate learning across different tasks.

5. A causal role for macaque area V2 in coarse disparity discrimination

5.1. Background

To perceive objects in 3-D space and enable us to act upon them, our visual system uses a variety of cues to determine depth. One important way through which this is achieved is by comparing small differences in the position on the retina of an object between both eyes, known as binocular disparity.

Neurons in many areas in visual cortex have been shown to be selective for binocular disparity, including in V1, V2, V3/V3A, MT, and V4 (Burkhalter & Van Essen, 1986; Poggio et al., 1988; Watanabe et al., 2002), amongst others (Gonzalez & Perez, 1998). Of these areas, V2 is the earliest stage in the visual hierarchy exhibiting a clear topographical map for disparity, with disparity-selective cells clustered by their preference, predominantly within cytochrome oxidase stained thick stripes (Chen et al., 2008; Ts'o et al., 2001). Similar disparity-selective topography has been observed in area MT (DeAngelis & Newsome, 1999b), as well as clustering in V3 (Adams & Zeki, 2001) and V4 (Watanabe et al., 2002). Given these areas all receive prominent direct and indirect input from V2 - in the case of MT directly from the thick stripes (DeYoe & Van Essen, 1985; Shipp & Zeki, 1985) - their organisation for disparity could be inherited from V2 (Chen et al., 2008; DeAngelis, 2000; Roe et al., 2007), positioning V2 as an important early stage in the perception of binocular disparity.

During a simple perceptual task, like disparity discrimination, the disparity sensitivity of single neurons in V2 is similar to the animal's behavioural sensitivity (Nienborg & Cumming, 2006), and is even correlated with the choices the animals made (Nienborg & Cumming, 2006, 2009). Similar results have been observed in other disparity-selective areas (Dodd et al., 2001; Uka, 2005; Uka & DeAngelis, 2004), one notable exception being V1, suggesting that a clear map for disparity may be a pre-requisite for these choice signals (Nienborg & Cumming, 2006). However, such correlations are unable to provide a causal link between these neurons and disparity perception, highlighted further by recent work finding decision-related activity in areas shown to not causally contribute to the decision (Yu & Gu, 2018).

Studies into the causal link between visual neurons and disparity perception have thus far focused on mid- and high-level visual areas (DeAngelis et al., 1998; Krug et al., 2013; Shiozaki et al., 2012; Uka & DeAngelis, 2006). These efforts have been fruitful, demonstrating a causal influence of neural activity in area MT for disparity perception (DeAngelis et al., 1998; Uka & DeAngelis, 2006) and motion from depth (Krug et al., 2013); and in area V4 for fine disparity discrimination, likely reliant on perception of relative disparities (Shiozaki et al., 2012). In contrast, the causal involvement of area V2, where we first see a topographic map for disparity, is still unclear.

5.2. Scientific Aims

In the present study, we aimed to address this prominent gap by investigating whether V2 causally contributes to disparity perception in a coarse disparity discrimination task. To do so, we used electrical microstimulation. The application of small alternating currents to selective clusters of neurons is thought to act as a boost to the signal of the stimulated neurons, akin to if the neurons had received greater sensory evidence in favour of their preferred task feature (Salzman et al., 1990). If the stimulated neurons are causally involved in perception, microstimulation would lead to a systematic increase in the number of choices towards the stimulated sites preferred feature vs the non-preferred feature (or “null” feature). Microstimulation has frequently been utilised as a tool to test causal relationships between a wide range of visual areas and perception of the features for which they are selective (e.g. Britten & van Wezel, 1998; DeAngelis et al., 1998; Salzman et al., 1990, see review from Cohen & Newsome, 2004). Additionally, the strength of microstimulation effects on perception has been found to depend on the tuning of the manipulated neurons, with a greater effect for those neurons with stronger selectivity (DeAngelis et al., 1998; Krug et al., 2013), as well as on non-sensory factors such as the expected reward on a given trial (Cicmil et al., 2015).

In this study we found that microstimulation of disparity-selective neurons in V2 led to a significant effect on behaviour in a coarse disparity discrimination task. Specifically, we saw an increase in the proportion of choices matching the preference of the stimulated sites. Further, we found that this effect was correlated with the strength of the neuron’s disparity tuning, and was mediated by reward size in one

animal, similar to findings in MT (Cicmil et al., 2015; DeAngelis et al., 1998; Krug et al., 2013). These results suggest that V2 is causally involved in absolute disparity perception, thus positioning it as one of the earliest visual areas to causally contribute to perception of binocular disparity.

5.3. Methods

5.3.1. *Animals*

We applied microstimulation to disparity-tuned units in visual area V2 in two male macaque monkeys (Macaca mulatta). Both animals were implanted with head fixation posts, and a recording chamber over the operculum of V1 under general anaesthesia. All procedures were in agreement with the Public Health Service policy on humane care and use of laboratory animals and were approved by the National Eye Institute Animal Care and Use Committee.

5.3.2. *Task*

The monkeys performed a coarse disparity discrimination task. After initiating fixation on a fixation point, a random-dot stereogram was presented (2 second duration). The animals' task was to report whether the central portion of this stimulus was protruding ('near') or receding ('far') with respect to its surround. After stimulus presentation, two choice targets appeared (one 'near' and one 'far') and the animals reported their choice with a saccade to the corresponding target.

5.3.3. *Reward Schedule*

Animals received a liquid reward after all correct trials (defined based on the stimulus) the size of which increased after consecutively correct trials. The maximum available reward size was awarded after 3 consecutively correct trials and remained high until the next error, after which it returned to the smallest available reward size.

5.3.4. *Stimuli*

Stimuli were generated on a Silicon Graphics (Mountain View, CA) workstation using custom-written software, and presented on two Flexscan F980 monitors (EIZO), each at 42cd/m2 mean luminance, 99% maximum contrast and 96Hz frame rate. The animals viewed the stimuli at a distance of 89cm through a set of two mirrors oriented at 45° and positioned 1.5cm from the eye (Wheatstone stereoscope configuration).

Stimuli were circular dynamic random-dot stereograms (RDS, 50% black, 50% white dots of 99% contrast, typically 0.09° radius, 40% dot density) similar to those described previously (Nienborg & Cumming, 2006, 2009). They consisted of a disparity varying center (2-5° size) and a surrounding annulus (1-2° width, shown at 0° disparity and always 100% correlated). Difficulty in the task was manipulated in one of two ways: either by varying the number of binocularly correlated dots (Nienborg & Cumming, 2006) (N=36) or the proportion of signal frames (Nienborg & Cumming, 2009) (N=11) in the center of each stimulus. Briefly, for both methods, one near and one far disparity were selected according to the tuning of the to-be-stimulated neurons (signal disparities). For the first method, on a given trial a percentage of dots in the central patch were either held at the near or far disparity, whilst the remaining dots were uncorrelated between eyes. For the second method, the dots in the central patch were 100% correlated, the disparity of which was varied frame-by-frame. On each trial signal disparity frames were interleaved with “noise” frames, the disparity of which was drawn from a uniform distribution of discrete, equally-spaced disparities, symmetric around 0° and encompassing the near and far signal disparities. Signal strength for the first method was determined by the the percentage of correlated dots, and for the second by the proportion of signal-to-noise frames.

5.3.5. *Microstimulation*

Prior to the task, disparity-selective sites in visual area V2 were identified for microstimulation. We recorded extracellular activity using single tungsten in glass electrodes (Alpha Omega, Nazareth, Israel). The signal was amplified (Bak Electronics, Mount Airy, MD), filtered (200Hz to 2kHz), digitised (32kHz), and stored to disk. Electrodes were inserted into V2 via the operculum of V1. V2 was identified as previously described (Nienborg & Cumming, 2006). Once in V2, disparity selectivity was measured in steps of 100um over 300um, and the site was selected as roughly the middle of a cluster with similar tuning preferences (similar to DeAngelis et al., 1998). The disparities that elicited near the highest and lowest responses were used as the preferred and null signal disparities respectively in the task. Stimulus size and position were determined based on the receptive field properties of the chosen site. Microstimulation was applied at 20uA (biphasic 200us cathodal followed by 200us anodal pulses at 200Hz) in randomly interleaved trials for the entire 2 second duration

of the stimulus presentation. Whether a choice was correct and rewarded was only determined by the visual stimulus.

5.3.6. *Analysis*

5.3.6.1. Inclusion criteria

Only sites with significant disparity tuning, determined using a one-way ANOVA ($p<0.05$), were used for further analysis. This led to the exclusion of one session. As the effects of microstimulation are known to decrease over time (e.g. Salzman et al., 1992), we restricted our analysis to include trials up to the first 80 complete microstimulation trials. This cut-off was similar to that used previously (Salzman et al., 1992). We additionally analysed the main effect of microstimulation over all trials for comparison, but the results did not qualitatively differ from those using only 80 trials.

5.3.6.2. Tuning Index

The strength of disparity selectivity for each session was calculated using the disparity discrimination index (DDI; see Prince et al., 2002) on the tuning curve collected prior to performing the task. This was calculated as

$$DDI = R_{max} – R_{min} / (R_{max} – R_{min} + 2RMS_{error})$$

where R_{max} and R_{min} are the highest and lowest square-rooted firing rates respectively, and RMS_{error} is the square-root of the residual variance around the means of square-rooted rates across all disparities.

5.3.6.3. Effect of Microstimulation

We measured the psychophysical performance of the animals as a function of the preferred tuning at each site. Trials were divided into two conditions depending on whether microstimulation was applied or not (control condition). Cumulative gaussians were fit to the behaviour for each condition in each session, and sessions with fits that explained less than 70% of the variance in either condition were excluded (N=34, 12 sessions excluded). The effect of microstimulation on behaviour was quantified as the shift (difference) in the point of subjective equality (PSE: defined as the mean of the cumulative gaussian). The same data was used to quantify the relationship between the PSE shift and tuning strength, which we did using a Pearson's correlation between PSE shift and DDI.

To quantify the main effect of microstimulation without being constrained by a strict variance explained criterion, we additionally collapsed trial data across all sessions (46 sessions) before making two single fits to the microstimulation and control conditions. Significance of the shift was determined using a permutation test, in which condition labels (microstimulation/control) were shuffled across sessions before measuring the PSE shift. This procedure was repeated 1000 times, and the size of the permuted shifts compared to the actual shift at a 5% significance threshold.

5.3.6.4. Effect of Reward Size on Microstimulation Effect

Trials on which the animal expected the maximum reward size were categorised as ‘large’ reward trials, and those where a sub-maximal reward was expected as ‘small’ reward trials (Cicmil et al., 2015). Trials were separated into four conditions in total, based on whether a large or small reward was expected, and whether microstimulation was applied or not. Cumulative gaussians were fit to the behaviour separately for each of these four conditions. We then quantified the effect of expected reward size on the size of the microstimulation-induced shift by computing the PSE shift separately for small and large reward conditions. Here, sessions with fits that explained less than 50% of the variance in any of the four conditions were excluded (N=29, 17 excluded). We used this relaxed variance explained criterion to allow the inclusion of a sufficient number of sessions which would allow us to estimate the effect (using 70% VE resulted in >70% of sessions being excluded). However, as described above, we additionally used a criterion-free approach to estimate the effect of reward size on the PSE shift for trials collapsed across all 46 sessions.

5.4. Results

Two macaque monkeys performed a coarse disparity discrimination task (Fig. 3). We presented a dynamic random-dot stereogram (RDS) for 2 seconds, and the animals had to decide whether the central portion of the RDS protruded ('near') or receded ('far') with respect to its surround. Two choice targets then appeared, corresponding to 'near' and 'far', and the animals indicated their choice with a saccade to the corresponding target. If the animals responded correctly, they received a liquid reward. The size of this reward increased with consecutively correct trials, and reached a maximum value after 3 trials. Whilst the animals performed the task, we applied electrical microstimulation to disparity-tuned clusters in V2 (N=47; Monkey A N=24; Monkey B N=23), identified prior to the task. 46 of these sites showed significant disparity tuning ($p<0.05$, one-way ANOVA, main effect of disparity) and were included in subsequent analysis. Microstimulation was applied on randomly interleaved trials for the entire stimulus presentation period. As the effect of microstimulation is known to decrease over time (Salzman et al., 1992), we restricted all main analyses to include up to the first 80 microstimulation trials.

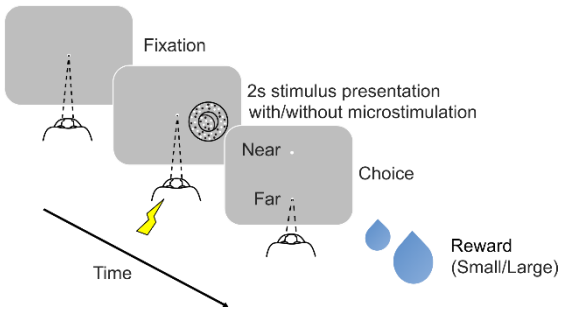

Fig. 3 | Task schematic. Animals were required to fixate at the centre of the screen. A random-dot stereogram was presented for 2 seconds, and the animals' task was to decide whether the central portion was "near" or "far" with respect to its surround. Microstimulation was applied on interleaved trials for the entire stimulus duration period. Two choice targets then appeared, signalling "near" and "far" choices, and the animal made his choice with a saccade to the corresponding target. If correct, they received a liquid reward, with larger rewards given for several consecutively correct trials.

5.4.1. *Effect of Microstimulation*

We first wanted to test whether microstimulation would induce a change in psychophysical behaviour. Specifically, whether stimulating disparity-selective neurons would lead to more choices towards the preferred or null disparity of those neurons, compared to when they were not stimulated. To address this, we split trials based on whether microstimulation was applied or not, and measured psychophysical performance as a function of the preference of the stimulated neurons. We then fit cumulative gaussians to the psychophysical performance in each condition, and excluded sessions where either fit explained less than 70% of the variance (N=34 after exclusion of 12 sessions).

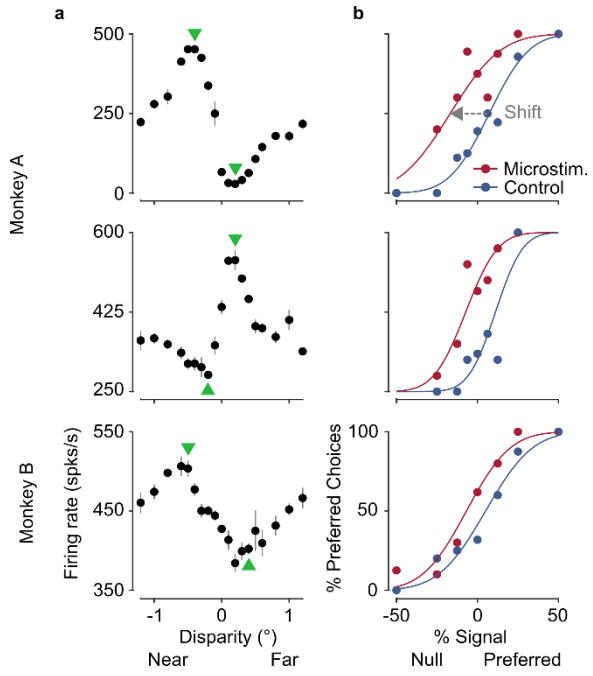

Fig. 4 | Example sessions. a) Disparity tuning for three example sessions of Monkey A (upper, middle) and Monkey B (lower). Two sessions are "near"-preferring (upper, lower) and one "far"-preferring (middle). Green triangles show the signal disparities chosen for the task. **b)** Psychophysical performance in microstimulation (red; line is cumulative gaussian fit, circles show raw data) and control (blue) trials, plotted with respect to the preferred disparity for each session (% preferred choices as function of preferred/null stimulus strength [% signal]). Dashed grey arrow (upper plot) illustrates the leftwards shift in the point of subjective equality (PSE) used to quantify the effect of microstimulation on behaviour.

Data from three example sessions clearly showed an effect of microstimulation on choices (Monkey A: Fig. 4a,b; Monkey B: Fig. 4c). In all three cases, the psychometric curve was shifted leftwards in the microstimulation condition, indicating an increase in the number of choices matching the preference of the stimulated neurons. In order to test whether this effect was consistent across the population, we quantified the microstimulation-induced shift of each individual site as the difference in the point of subjective equality (PSE) between conditions (control – microstimulation).

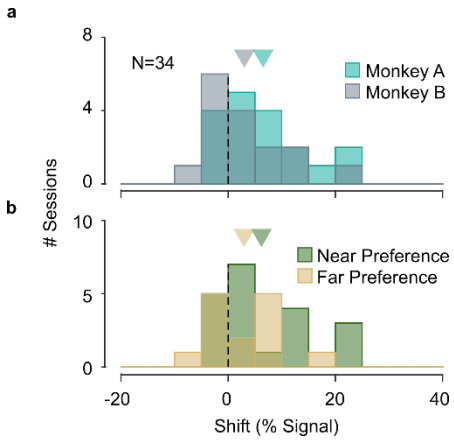

Fig. 5 | Microstimulation leads to a shift in the psychometric function. PSE shifts (% signal, control - microstimulation) for sessions with fits explaining >70% variance, split by **a)** monkey, and **b)** disparity preference.

Indeed, there was a significant leftwards shift of the psychometric function (Fig. 5, N=34, mean shift (% signal)=4.87, p=0.002, two-tailed t-test). Furthermore, there was no difference in the distribution of shifts between animals (monkey A: N=18, mean=6.51; monkey B: N=16, mean=3.02; difference, p=0.22 , two-sample t-test) or between sites with different disparity preferences (near: n=20, mean=6.20; far: n=14, mean=2.96 ; difference, p=0.26 , two-sample t-test). Thus, microstimulation in V2 indeed biased choices towards the preferred disparity of the stimulated cells.

We additionally used a variance explained (VE) criterion-free approach to measuring the significance of this effect, to ensure our result was not influenced by the specific VE criterion we used. To this end, we combined trials across all significantly tuned sessions, separately for each monkey (monkey A: n=24, monkey B: n=22). We then fit cumulative gaussians to the combined data from the microstimulation and control conditions of each monkey (Fig. 8a,b). In both animals, we again observed a significant leftwards shift of the psychometric function (monkey A, shift=5.72, p=0.015; monkey B, shift=5.38, p=0.029; two-tailed permutation tests).

Having shown evidence of a causal effect for V2 in disparity choices, we next sought to characterise further aspects of this relationship. Experiments using microstimulation have previously established evidence of a causal relationship between area MT and disparity discrimination during similar tasks (Cicmil et al., 2015;

DeAngelis et al., 1998; Krug et al., 2013). We thus aimed to investigate whether the same principles governing the effect in MT hold for V2.

5.4.2. *Relationship with tuning*

Motivated by previous results from area MT (DeAngelis et al., 1998; Krug et al., 2013), we asked whether the size of the microstimulation-induced shift was influenced by the tuning strength of the stimulated sites in V2. To answer this question, we measured the tuning strength of each significantly tuned site using the disparity discrimination index (DDI) and compared this with the PSE shifts from sessions with fits that explained >70% of the variance (Fig. 6). The range of DDI values was small (0.54-0.88); nonetheless, we found a significant positive correlation between DDI and PSE shift (n=34, r=0.51, p=0.002; Monkey A: N=18, r=0.55, p=0.019; Monkey B: N=16, r=0.51, p=0.045; All Pearson correlations). The effect of microstimulation on choices was thus greater for sites with stronger disparity tuning.

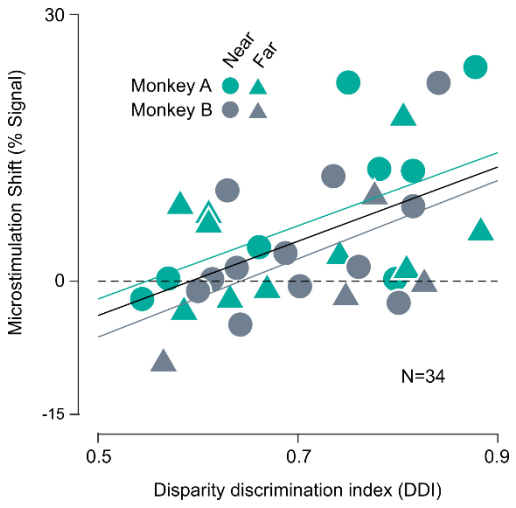

Fig. 6 | Size of microstimulation effect is correlated with tuning strength. Microstimulation-induced shifts in the PSE (% signal) as a function of tuning strength, as measured by the disparity discrimination index (DDI), for sessions with fits explaining >70% of the variance. Data is shown for monkey A (green) and monkey B (grey), for near (circle) and far (triangle) disparity preferences. Lines show linear regression for both animals (black), monkey A (green) and monkey B (grey).

5.4.3. *Effect of reward size on microstimulation*

Expected reward size has previously been found to influence the strength of microstimulation effects in area MT during a perceptual discrimination task (Cicmil et al., 2015). Using a motion discrimination task that also required depth judgements (structure-from-motion cylinders), the authors found the size of the microstimulation-induced PSE shift to be smaller before large than before small expected rewards. This was suggested to reflect an influence of expected reward on decision-making as early as the sensory processing level.

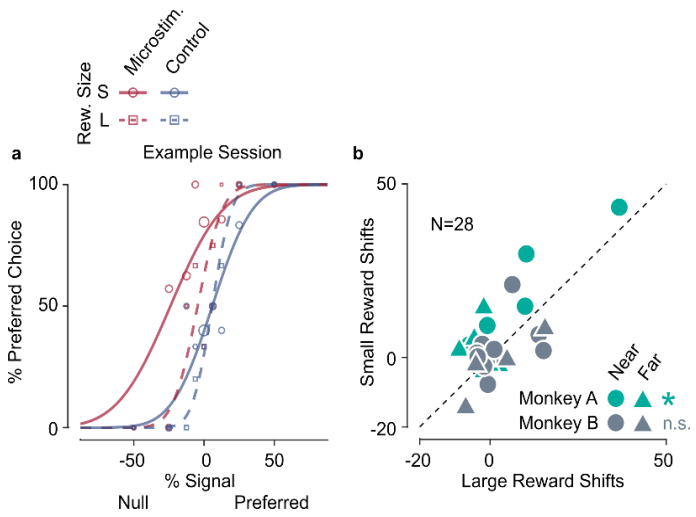

Fig. 7 | Size of microstimulation effect depends on expected reward size in one animal. a) Psychophysical performance (% preferred choices as function of preferred/null stimulus strength [% signal]) for one example session from monkey A in microstimulation (red) and control (blue) trials with a small (solid lines, cumulative gaussian fits; circles, size proportion to number of trials) or large (dashed lines; squares) expected reward. **b)** Microstimulation-induced shifts of the PSE for large (x-axis) and small (y-axis) expected rewards, for sessions with fits that explained >50% of the variance. Data shown for monkey A (green) and monkey B (grey), for near (circles) and far (triangles) disparity preferences. Asterisk denotes significance.

To test whether microstimulation in area V2 during coarse disparity discrimination showed a similar dependence on reward size, we divided trials into two groups, based on whether a large or small reward was expected. Within each group, we further split the data into microstimulation and control trials. We fit cumulative gaussians separately to each of the four conditions, and excluded sessions for which any of the fits explained less than 50% of the variance (see Methods for exclusion criteria). We then measured the microstimulation-induced PSE shift separately for the large and small reward conditions.

Data from an example site from monkey A (Fig. 7a) showed a clear and large effect of microstimulation on performance for small reward trials (shift [% signal]=29.84), but a much smaller effect for large reward trials (shift=10.29). This effect was consistent across sessions in monkey A (Fig. 7b, green) with significantly smaller PSE shifts before large expected rewards compared to small expected rewards (N=13, small mean=9.16, large mean=2.09, p=0.004). However, we did not find the same effect for monkey B (Fig. 7b, grey; N=15, small mean=1.15, large mean=2.27,

p=0.55), and the effect was not significant when combining data across both animals (n=28, small mean=4.87, large mean=2.19, p=0.09, paired t-test). As for the main microstimulation analysis above, we additionally used a VE criterion-free approach for our reward size analysis, to test the effect of reward size on microstimulation independently of a specific criterion or the exclusion of a large number of sessions. We collapsed trials across all significantly tuned sessions, before splitting them based on reward size and whether microstimulation was applied. We fit a cumulative gaussian to each of the four conditions, and measured the PSE shift for large and small reward trials separately. We again found a significantly smaller microstimulation-induced shift in large reward than in small reward trials for monkey A (Fig. 8c, n=24, small

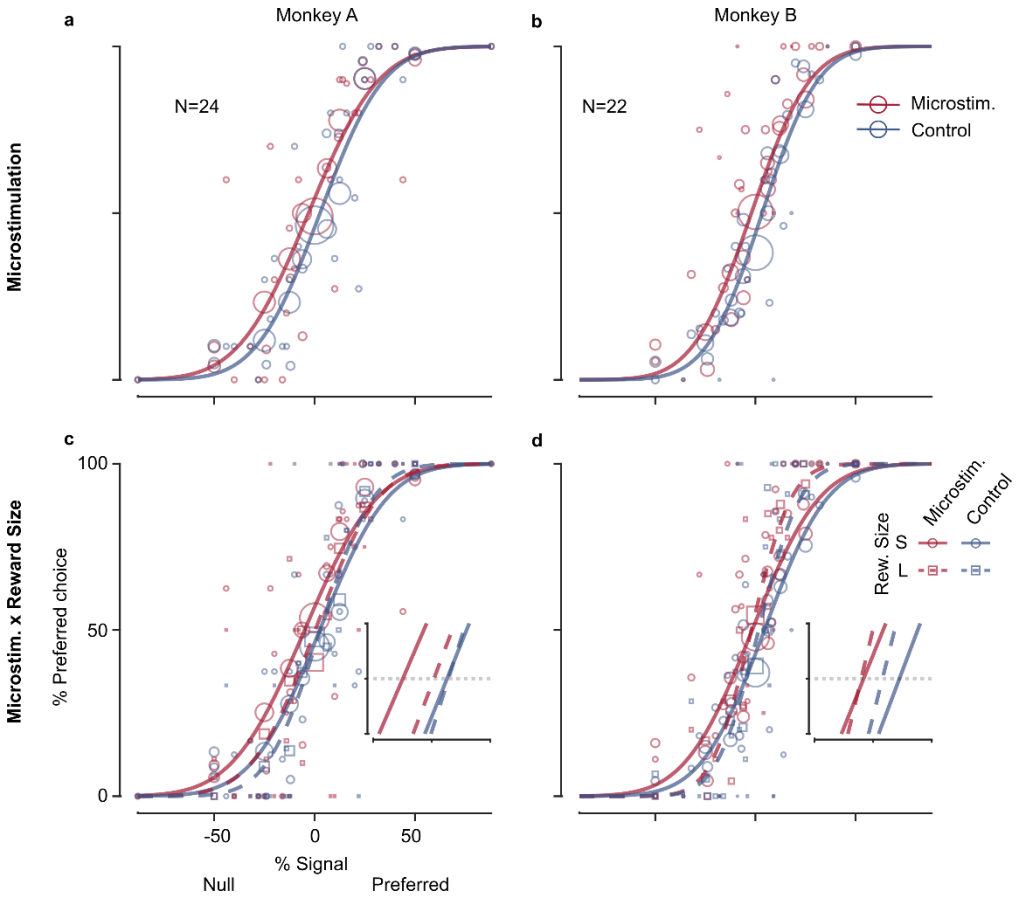

Fig. 8 | Microstimulation and reward size results remain unchanged without variance explained criterion. Psychophysical performance (% preferred choices as function of preferred/null stimulus strength [% signal]) collapsed across all sessions for monkeys A (left) and B (right). **a,b)** Performance plotted for microstimulation (red, cumulative gaussian fit) and control (blue) trials (circles, size proportion to the number of trials per session). **c,d)** Same, but additionally split by trials with small (solid lines, cumulative gaussian fits; circles, size proportion to number of trials) or large (dashed lines; squares) expected reward. Insets show fits around the 50% preferred choice mark (dashed line).

mean=7.54, large mean=2.48, p=0.029; two-tailed permutation test), but not for monkey B (Fig. 8d, n=22, small mean=6.11, large mean=3.63, p=0.54; two-tailed permutation test).

5.5. Discussion

In sum, we found that microstimulation of neurons in area V2 systematically biased choices towards the disparity preference of the stimulated neurons in a coarse disparity discrimination task. This suggests a causal role for V2 in the perception of binocular disparity. Importantly, V2 is the earliest visual area for which such evidence has been shown, suggesting it may play an important role as an early processing stage in the perception of binocular disparity.

5.5.1. *A role for V2 in coarse disparity discrimination*

Our findings specifically show a causal role for V2 in coarse disparity discrimination. This task is thought to rely on the perception of absolute disparities, that is the absolute difference in the position of an object between both eyes. V2 is also the first area known to exhibit some selectivity for small differences between absolute disparities (Thomas et al., 2002), known as relative disparity. The perception of relative disparities is believed to underlie the ability to perform fine disparity discriminations for which areas in the ventral stream are thought to be responsible (e.g. V4 (Umeda et al., 2007) and IT (Uka, 2005), as well as V2 (Clery et al., 2017)). Given V2's position near the start of the ventral and dorsal streams, V2 may be involved in processing both absolute and relative disparities before separately relaying each type of information to its relevant stream. Whilst this is not discernible from our findings, combining causal manipulation of V2 during fine and coarse disparity discrimination tasks with recordings in mid to late visual areas could help shed more light on the role of V2 in the disparity processing hierarchy.

5.5.2. *Comparison with MT*

We additionally showed that the overall effect of microstimulation in V2 was modulated by several factors, as has been previously demonstrated in area MT during similar tasks. This opens the question of the relationship between both areas with respect to the processing of disparity. Firstly, MT has also exhibited a causal role during coarse disparity discrimination. It is plausible that V2 and MT constitute different stages in the

same processing hierarchy. This is supported by anatomical evidence showing that MT receives disparity-selective input directly from the thick stripes of V2 (DeYoe & Van Essen, 1985; Shipp & Zeki, 1985). In addition, when V2 and V3 are inactivated, MT has been shown to lose much of its selectivity for disparity (Ponce et al., 2008). While MT's causal involvement in motion discrimination (Salzman et al., 1992; Salzman et al., 1990), and perceptual tasks involving conjunctions of motion and depth (Krug et al., 2013), sets it apart as an integral processing stage in such tasks, it is unclear what additional role MT affords in a depth-only discrimination task like the one we present here.

Secondly, we also found that the effect of microstimulation was correlated with the degree of tuning of our V2 neurons, as in MT (DeAngelis et al., 1998; Krug et al., 2013). This suggests that there may be a relationship between the task-relevance of a neuron, i.e. how well it is tuned for the discriminated features, and the influence it exerts on the decision. Again, it is possible that MT inherits this relationship from V2. Alternatively, it may reflect some interaction between V2 and MT that leads to a cumulative strengthening of the readout of the most informative neurons for the task.

However, it is also possible that these common findings reflect at least partially parallel processes. Future work on this topic could seek to better characterise the relationship between V2 and MT with regards to disparity processing, as well as to isolate possible unique contributions to performance in such a task.

5.5.3. *The role of reward size in decisions*

Lastly, in one animal we also found an effect of expected reward size on microstimulation in V2, such that larger rewards led to a smaller effect of microstimulation on disparity perception. Overall, this lends further support to the claim that expected reward can affect the decision-process as early as the sensory stage. However, since in our data this effect was only present in one animal, the influence of reward on the decision-process is not so clear-cut. As reward is also known to influence decision-making at a later evidence integration stage (Rorie et al., 2010), it is plausible that the asymmetry between our animals was driven by a difference in the magnitude of the effects of reward at earlier and later stages of the process. This would in turn affect the interaction with the microstimulation effect. Whether this could arise as a

result of differences in perceptual strategy or other inter -individual differences is unclear and requires future work.

Overall these results not only establish an important role for V2 neurons in disparity perception, but also provide exciting new avenues for future research.

6. Decision-related feedback in visual cortex lacks spatial selectivity

6.1. Background

The brain excels at flexibly performing a multitude of tasks. This ability likely requires the relevant neuronal circuits to have access to task-relevant contextual information. The communication of such context could be supported through feedback signals to upstream populations in accordance with task demands (Harris & Mrsic-Flogel, 2013; Roelfsema & de Lange, 2016). In the visual system, such feedback has been implicated in the modulation of neurons representing task-relevant variables (Bondy et al., 2018; Cohen & Newsome, 2008; Haefner et al., 2016; Roelfsema & de Lange, 2016). Current thinking suggests that feedback to the visual cortex mediates context-dependent predictions (Rao & Ballard, 1999), perceptual learning (Roelfsema et al., 2010), beliefs for hierarchical Bayesian inference (Haefner et al., 2016; Lee & Mumford, 2003), expectations (Coull et al., 2000; Summerfield & de Lange, 2014), gating (Moore & Armstrong, 2003) or tagging of the relevant sensory information (Haimerl et al., 2019, 2021) to support downstream processing.

These accounts predict the feedback to be selective, targeting some sensory information over other depending on the context of the task and stimulus (Bijanzadeh et al., 2018). Anatomical evidence supports some selectivity of the feedback connections in the visual system (Briggs, 2020; Federer et al., 2021) but the extent to which this enables the selective targeting of specific subsets of neurons is unknown. In light of the enormous number of ethologically possible tasks and contexts, such task-dependent selectivity could become anatomically costly. Limiting the selectivity of the feedback may also be beneficial by facilitating generalization across different tasks.

6.2. Scientific Aims

Here, we set out to test how flexible the selectivity of this feedback is. Specifically, we explored whether previously reported task-specific modulation by feedback (Bondy et al., 2018) is selective for neurons representing a relevant stimulus in the presence of task-irrelevant stimuli. We extended a widely used visual discrimination paradigm (Britten et al., 1996) to include both task-relevant and task-irrelevant stimuli at different

spatial locations; thus performance in this task required spatial selectivity in addition to selectivity for visual discrimination. During simple visual discrimination tasks using a single stimulus, visual neurons are typically correlated with an animal's choice, unexplained by the stimulus ("choice-correlations") (Britten et al., 1996; Pitkow et al., 2015). Previous work identified a significant decision-related feedback component of these choice-correlations (Bondy et al., 2018; Nienborg & Cumming, 2009; Wimmer et al., 2015). Conversely, tasks directing attention to one spatial location over others have identified spatially selective modulation of responses in the visual cortex (Moran & Desimone, 1985).

In the current study, performance in the task in principle required both aspects of selectivity. Thus, we can measure the spatial selectivity of the decision-related modulation by feedback to see how the additional task demands shaped the feedback. If the decision-related feedback modulates the visual neurons selectively according to their task-relevance, it should not affect neurons representing a task-irrelevant stimulus, and these neurons should hence not be correlated with choice (Fig. 9b, c). This also should be the case if choice-correlations only reflected feed-forward effects, cf. (Shadlen et al., 1996), in which case the predictions for choice-correlations would be identical to those for the selective feedback, Fig. 9e, f. Conversely, if the decision-related feedback is unselective to whether the neurons representing the stimulus are relevant for the task, the neurons should show correlations with choice even when representing a task-irrelevant stimulus (Fig. 9h, i). In fact, it has been hypothesized that decision-related feedback engages the same neural mechanism as feature-based attention (Bondy et al., 2018; Cohen & Newsome, 2008; Krug, 2004). Studies examining feature-based attention showed that when a subject's attention was directed to a particular stimulus feature, the response of neurons selective for this feature was increased (Bichot et al., 2005; Cohen & Maunsell, 2011; Treue & Trujillo, 1999). A defining characteristic of such modulation by feature-based attention is that it is observed throughout the visual field (Saenz et al., 2002; Treue & Trujillo, 1999; Wojciulik et al., 1998). As a consequence, if decision-related feedback is linked to the spatially global feedback of feature-based attention, this predicts that neurons representing a task-irrelevant stimulus are correlated with choice (Fig. 9h, i).

Here, we show that although the animals' behavior is highly spatially selective, the decision-related feedback is not. The lack of selectivity cannot be explained by

stimulus effects, behavioral covariates or stimulus- and task-independent neuronal covariability (“noise-correlations”). At the level of simultaneously recorded populations, the representation of choice and stimulus is partially misaligned. These stimulus-choice (mis)alignments are similar whether the stimulus is relevant or not. Our results support the previously hypothesized link between feature-based attention and decision-related activity and reveal a feed-back mechanism that may support generalization across tasks.

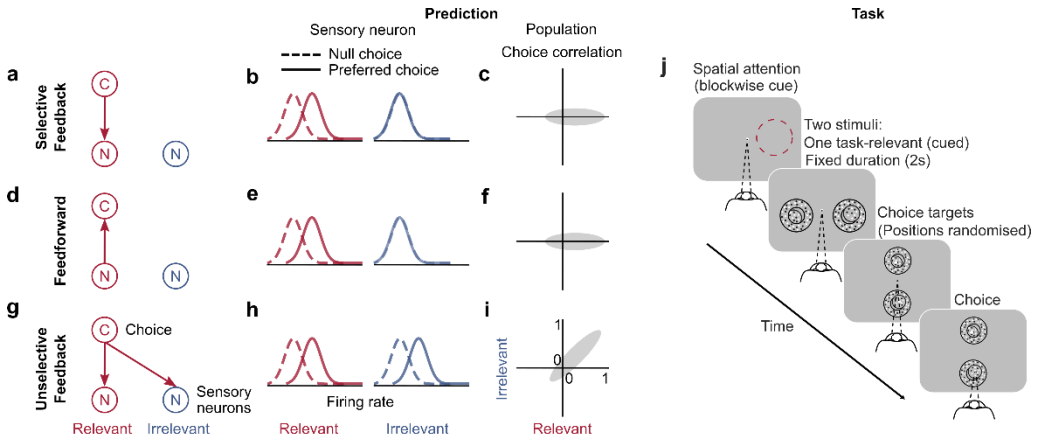

Fig. 9 | Predictions and task. (**a**-**i**) Predictions for choice-correlations of visual neurons representing a task-relevant or an ignored, task-irrelevant stimulus. Left: Neurons representing relevant (red) and irrelevant (blue) sensory information (N), or choices (C). Middle: Predicted firing rate distributions of sensory neurons representing relevant (red) or irrelevant (blue) stimuli, separated for trials on which the observer chose a neuron’s preferred (solid) or null stimulus (dashed). Right: Predicted distribution of choice-correlations across a population of sensory neurons representing the relevant (abscissa) or irrelevant (ordinate) stimulus. For selective feedback (a-c) or feedforward processes (d-f) choice-correlations are only expected for neurons representing the task-relevant stimulus. If feedback is unselective choice-correlations are also expected when neurons represent the irrelevant stimulus (g, h, i). For feature-discrimination tasks, spatially global feature-based attention predicts such unselective modulation if these feed-back processes are linked. (**j**) Monkeys performed a disparity discrimination task on a cued stimulus while a task-irrelevant stimulus was simultaneously presented in the opposite hemifield. The relevant hemifield (red circle only for schematic) was indicated at the beginning of each block.

6.3. Methods

6.3.1. *Animals*

Two adult male rhesus monkeys (*Macaca mulatta*; A, 7kg; B, 9kg, both 13 years old; housed in pairs) were implanted with a titanium head-post and two titanium chambers

over the operculum of V1 in both hemispheres under general anesthesia. All experimental procedures were approved by the relevant local authority, the Regierungspräsidium Tübingen, Germany.

6.3.2. *Behavioral task*

The monkeys performed a coarse disparity discrimination task on one of two stimuli presented on the screen (Fig. 9j). The task-relevant hemifield was cued at the beginning of each block (50 trials) by three trials during which a single stimulus was presented on the task-relevant side. Once the animal fixated on a fixation point, two dynamic random-dot stereograms (each analogous to Nienborg & Cumming, 2009) were presented simultaneously (2 sec duration), one in each hemifield. Both stimuli were statistically identical but independently varied, and only the task-relevant stimulus was informative about the correct choice. The task was to report whether the central disk of the cued stimulus was protruding (“near”) or receding (“far”) with respect to its surround. After the stimulus presentation two choice icons (one indicating a “near”, one a “far” choice, both typically horizontally offset towards the cued side by ~1-3°) appeared, whose vertical position (typically 3°- 4° above and below the fixation point, held constant within a session) was randomized from trial to trial. This ensured that during the stimulus presentation the motor-command-to-choice mapping was unknown to the animal. The monkeys reported their decision with a saccade to one of the choice icons. Correct choices were rewarded (Cicmil et al., 2015) with a liquid reward.

6.3.3. *Electrophysiological recordings*

We recorded extracellular single and multi-unit activity in areas V2 and V3/V3a using multi-channel laminar probes (Plexon, TX, USA; V/S Probes, 24/32 channels, 50-100um inter-contact spacing). Eye movements were tracked binocularly using the Eyelink 1000 (SR Research, OTT, Canada) at a sampling frequency of 500Hz. Neuronal signals were amplified, digitized, and filtered (250 Hz to 5 kHz) with the Ripple Grapevine System (Ripple Neuro, UT, USA).

6.3.4. *Procedure*

Recording sites were initially mapped using single tungsten in glass electrodes (Alpha Omega, Nazareth, Israel), and selected based on their disparity selectivity and receptive field position.

Probes were inserted into V2 and/or V3/V3a via the operculum of V1, approximately orthonormal to the surface, guided by anatomical MRI scans of the animals' brains, using a microdrive system (NaN Instruments, Israel) with custom-made mounts. V2 was identified as previously described (Nienborg & Cumming, 2006) and verified offline based on a shift in receptive field position and size compared to those in V1. After characterization of the V2 receptive field positions the probes were advanced further and V3/V3a was identified on the basis of shifted and larger receptive fields with respect to those in V2 and clustering for binocular disparity, cf. (Hubel et al., 2015). Given the previously reported similarity between the disparity selectivity in V3 and V3a (Anzai et al., 2011) we collapsed across recordings in V3 and V3a.

6.3.5. *Stimuli*

Visual stimuli were back-projected on a screen using a DLP LED Propixx projector (VPixx, Saint-Bruno, Canada; 1920 X 1080 pixels resolution; 30cd/m² mean luminance; linearized gray values; run at 100Hz or 60Hz for each eye) combined with an active circular polarizer (DepthQ, Lightspeed Design Inc., WA, USA; run at 200Hz or 120Hz). The monkeys viewed the screen (viewing distance: 103.0 and 97.5 cm in monkeys A and B, respectively) binocularly through passive circular polarizing filters. Visual stimuli were generated in Matlab (Mathworks) using custom written code based on Eastman & Huk (2012) using the Psychophysics toolbox (Kleiner et al., 2007).

Stimuli were circular dynamic random-dot stereograms (RDS, 50% black, 50% white dots, typically 0.08° radius, 50% dot density). They consisted of a disparity varying center (updated at the framerate of the display, i.e. at 100Hz or 60Hz) and a surrounding annulus (1° width, shown at 0° disparity), the size and position of which was determined by the aggregate receptive field of the recorded units (mean RDS center size: 3.6°; mean stimulus eccentricity: 6.3°). On a given stimulus frame, the center dots all had the same disparity, but the disparity could change from frame to frame. Signal disparities (always one near and one far disparity value in each session)

were selected to approximately match the disparity selectivity of the majority of the recorded units. Signal frames were interleaved randomly with “noise” frames. The disparity of these noise frames was drawn from a uniform distribution of typically 9 values of discrete, equally-spaced disparities (symmetrical about 0° disparity, typical values were -0.4°,-0.3°, -0.2°, -0.1°, 0°, 0.1°, 0.2°, 0.3°, 0.4°), encompassing the near and far signal disparity. Signal strength (measured in % signal, signed, where negative values refer to near signal trials, and positive to far signal trials) in a given trial was determined by the proportion of signal to noise frames, and was used to manipulate task-difficulty. For example, -10% signal refers to a trial on which for 10% of the stimulus frames (randomized over time) the central region of the stimulus had the near signal disparity, while the disparity on the remaining 90% of the frames was drawn from the noise distribution. On 0% signal (“no-signal”) trials, i.e. all frames were drawn from the noise distribution, the correct choice was undefined, and the animal was rewarded randomly on 50% of the trials. The target icons were also RDS but slightly smaller than the stimuli, and always presented at 100% near and far signal.

In a subset of sessions, the random seed used to generate the stimuli was fixed on half of the trials to produce identical stimuli (“frozen noise”, see Fig. 16).

Disparity tuning curves were measured prior to the behavioral task using identical RDS as used for the task but shown for 450ms each at 100% signal at changing disparities (typically -1° to 1° in 0.1° increments).

For the control experiments (Fig. 17), identical RDS stimuli as for the task were used with fixed random seeds for a range of signal disparities encompassing the preferred and null disparity of the recorded units.

6.3.6. *Analysis*

Single and multi-unit activity (collectively referred to as units) was sorted offline using the Plexon Offline Sorter (v3.3.5).

6.3.7. *Inclusion criteria*

Only successfully completed trials were included for further analysis. For each unit periods of pronounced non-stationarity were removed. To do so we computed a running 20-trial average spike-count, calculated separately for each attention

condition. Periods for which this running average decreased below 20% its peak were removed, and the longest continuous segment of included trials was used for further analysis. This resulted in the removal of 8 (0.6%) of the V2 units and 24 (2%) of the V3/V3a units because the remaining data after removing periods of non-stationarity did not meet the minimum number of trials to compute choice-correlations (at least 5 trials with near and far choices, respectively for the 0% signal stimulus). Additionally, only units for which the mean response to the 0% signal stimulus exceeded 4 spikes/sec, and for which the d-prime to discriminate the stimuli with the highest signal strength was >0.5 were included. Additionally, the following behavioral criteria had to be met for each unit: the animal's performance had to exceed 70% for the highest signal stimuli and the bias for 0% signal trials had to be below 75%. For the V2 dataset, out of 1301 units 152 (12%), 266 (20%) and 172 (13%) were excluded because the criteria for minimum firing rate, d-prime or behavior, respectively, were not met. For V3/V3a, out of 1035 units, 123 (12%), 298 (29%) and 104 (10%) were removed for these reasons, respectively. Of the included 703 V2 (486 V3/V3a) units 5.5% (3.5%) of trials were excluded due to non-stationarities. Of the included units 18/703 (15/486) were single units in V2 (V3/V3a). The main findings for the single units were qualitatively similar to those for the multi-unit activity ($p>0.35$ for all Wilcoxon rank sum tests on differences in medians for the choice-correlations for the relevant and irrelevant stimulus in either area and across areas), and we therefore collapsed all analyses across single-units and multi-units.

6.3.8. *Receptive field positions*

For each unit we measured a horizontal and vertical response profile (as described in Seillier et al., 2017). These reflect the one-dimensional receptive fields examined along the horizontal axis (using an elongated vertical stimulus, typically a low-spatial frequency sinusoidal luminance grating 0.3-0.5 by 3-5 degrees) or vertical axis (using a horizontally elongated stimulus). These were fit separately with Gaussian functions, and fits were required to explain at least 70% of the variance. The average receptive

field position for each area and session (Fig. 10) was computed as the average mean of all included fits to the horizontal (x) and vertical (y) response profiles.

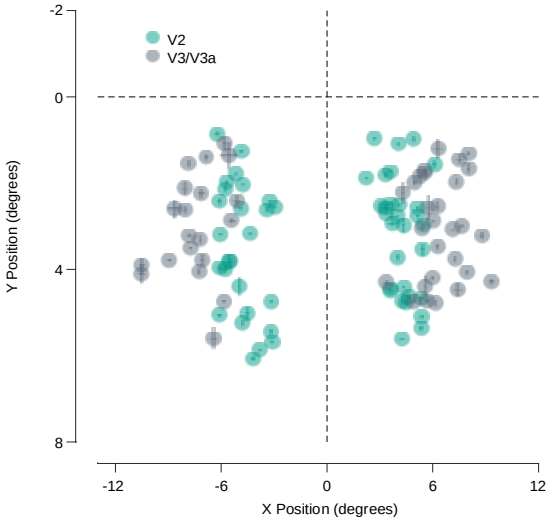

Fig. 10 | Receptive field centers. Average receptive field positions (in °) are shown for each included session for V2 (green) and V3/V3a (gray) and monkeys A and B. For each penetration, horizontal and vertical receptive field profiles were measured from the responses on each channel using an elongated grating or bar, and fit by Gaussian functions. The average center and width of the receptive fields was computed from the mean and SD of fits which explained >70% of the variance (V2 mean horizontal center=4.48°, width=±1.04°, vertical center=3.38°, width=±0.95°; V3/V3a mean horizontal center=6.83°, width=±1.54°, vertical center.

6.3.9. *Behaviour*

Performance was measured as percent far choices as a function of the cued or uncued stimulus' signed signal strength. Cumulative Gaussians were fit to the session-averaged performance for the cued stimulus. Psychophysical thresholds and bias were defined as the standard deviation and mean of these cumulative Gaussians.

6.3.10. *Psychophysical reverse correlation*

Time-resolved psychophysical kernels (Neri et al., 1999) were computed for 0% signal trials (see also Kawaguchi et al., 2018; Nienborg & Cumming, 2009) for the cued and uncued stimuli separately. For each of four non-overlapping consecutive time bins (500ms each), the stimulus was converted to an n-by-m matrix (n, number of discrete disparity values used for the stimulus; m, number of trials). Each entry of this matrix

contained the proportion of frames on which a given disparity was presented in this time-bin and trial. The kernel for each time-bin was computed as the difference between the mean matrix across near-choice trials and the mean matrix across far-choice trials. The kernels were averaged across sessions, adjusted for the frame-rate and weighted by the number of trials per session.

To further confirm that the animals relied correctly on the relevant rather than the irrelevant stimulus we also restricted the analysis of the psychophysical kernels to those trials for which both the relevant and irrelevant stimulus had no net disparity signal (no-signal stimulus). Note that because of the reduced number of trials these kernels were noisier but the structure mirrors remained unchanged.

6.3.11. *Generalized linear model analysis of behavior*

To examine the degree to which the relevant or irrelevant stimulus or their interaction influenced the animals' choices (a far choice is defined as 1) we fit a generalized linear model (GLM) to the animals' behavior in each session:

$$(1) \quad P(choice = 1) = \Phi(\beta_0 + \beta_1 s_{rel} + \beta_2 s_{irrel} + \beta_3 s_{rel} s_{irrel})$$

where $\Phi()$ is the cumulative distribution function of the normal distribution, s_{rel}, s_{irrel} and $s_{rel}s_{irrel}$ correspond to the relevant and irrelevant stimulus on each trial (in signed percept signal strength) and their interaction, respectively. We fit weights (β_1, β_2, β_3) to account for the contribution to the animals' choices of each of these covariates as well as the animals' bias (β_0). To ensure comparability of the resulting weights we normalized (z-scored) the covariates prior to fitting. We used lasso regularization (function *lassoglm* in Matlab; 10-fold crossvalidation; (Hastie et al., 2009) to avoid overfitting.

6.3.12. *Modulation by spatial attention*

To measure the modulation by spatial attention we compared the mean responses (*R,* in spikes/sec during the 2sec stimulus presentation) to the 0% signal stimulus when it was cued versus uncued. We quantified the modulation by a contrast (cf. Treue & Trujillo, 1999) as a spatial attention index, $AI = (R_{cued} - R_{uncued})/(R_{cued} + R_{uncued})$.

6.3.13. *Choice-correlations*

Choice probabilities (signed according to the tuning for disparity of each unit, measured during the fixation tasks outside of the discrimination task) were computed based on the average firing rates for the 0% signal trials. The mean responses, corrected for stimulus induced effects as described below, for each trial were grouped according to the choice the animal made on a that trial. From the distribution of firing rates when the animal chose a neuron's preferred and null disparity, we computed the area under the receiver operating curve (aROC), which was defined as choice probability (Britten et al., 1996). Choice probabilities (*CP*) can be converted to a Pearson's correlation coefficient "choice-correlations", c_{choice}, between the neuronal response and a continuous decision variable as derived previously, i.e. eq. S1.7 in Haefner et al. (2013) and eq. 8 in Pitkow et al. (2015):

$$(2) \quad CP = \frac{1}{2} + \frac{2}{\pi}\tan^{-1}\frac{c_{choice}}{\sqrt{2 - {c_{choice}}^2}}$$

Based on this relationship we converted choice probabilities to choice-correlations as done in Lueckmann et al. (2018). Note that Pitkow et al. (2015) used a linear approximation to this quantity: $CP \approx \frac{1}{2} + \frac{\sqrt{2}}{\pi}c_{choice} \approx 0.45c_{choice} + 0.5$. Because we define the preferred and null disparity based on each unit's disparity tuning, choice-correlations are signed with respect to each unit's tuning. That is, positive c_{choice} values mean that a unit has a higher firing rate on trials when the animal chooses a unit's preferred disparity. Conversely, negative c_{choice} values imply lower firing rates on trials when the animal chooses the unit's preferred disparity.

Before computing choice-correlations the mean responses for each trial were corrected for stimulus-induced effects (Nienborg & Cumming, 2009). Psychophysical reverse correlation takes advantage of systematic differences in the stimuli preceding the monkeys' choices. The psychophysical kernel demonstrates that there are systematic differences in the stimuli preceding the monkeys' choices even for the stimuli that on average contained 0% signal. Choice-correlations were therefore corrected for this stimulus-induced component. This correction was done in two steps. First, we extracted a subspace map (for all 0% signal trials, not separated by choice), s (Nienborg & Cumming, 2009). The subspace map s is an n-dimensional vector giving

the total number of spikes (si) elicited by one frame of a given disparity di. We then summarized the stimulus for each trial by an n- dimensional vector (u). This vector u represents the number of frames with which each disparity occurred on a given trial. The number of dimensions (n) corresponds to the number of disparities included in the experiment. For each trial, the inner product between s (number of spikes per one frame of disparity di) and the stimulus vector u (number of frames per disparity) was calculated. The resulting spike count is the spike count predicted based on the disparities included in a given trial and the subspace map. It varies systematically from trial to trial, depending on u, i.e. the actually included disparity-signals on a given trial. (Although, on average the distribution was flat.) The resulting predicted spike count was therefore subtracted from the actually measured spike count on that trial. We note that this correction does not remove random fluctuations in firing as a function of choice.

The time-courses of the choice- correlations were computed analogously but for the firing rate in a sliding (10ms steps) 300ms wide window, and corrected for the latency-adjusted stimulus-induced effect (Nienborg & Cumming, 2009).

6.3.14. *Choice-correlations for "frozen noise" stimuli*

To quantify choice-correlations we corrected stimuli for the random fluctuations of the white-noise stimuli used to perform psychophysical reverse correlation. To verify that this correction was adequate we computed choice-correlations for repeated presentations of the same random sequence of disparity values ("frozen noise") in a subset of sessions. We note that to allow for the psychophysical reverse correlation analysis to verify the spatial selectivity of the behavior, we needed a subset of random noise stimuli even for those sessions in which we included frozen noise stimuli. We therefore fixed the random seeds on half of the trials and randomized them on the other half. But for this reason the number of trials for which choice-correlations were computed was reduced to 50%, and made our estimates of the choice correlations noisier.

6.3.15. *Eye movement controls*

6.3.15.1. *Effects of mean eye position*

Given the randomization of the position of the choice targets an association between the animals' eye-position during the stimulus presentation and choice was unlikely. However, to ensure this was the case we checked for any differences in eye positions preceding a target presented above or below the fixation marker, preceding a saccade to the upper or lower target, and preceding near or far choices.

To verify that any differences in these measures could not account for observed choice-correlations we then measured the effects of mean vertical or horizontal eye position on firing rate within each choice and then regressed out the predicted effect across choices.

6.3.15.2. *Effects of microsaccades*

Microsaccades have been shown to weakly modulate the activity of visual neurons (e.g. Harris & Mrsic-Flogel, 2013; Roelfsema & de Lange, 2016). We therefore explored any systematic relationship between the frequency, amplitude and direction of microsaccades and choice-correlations. Microsaccades were labeled using a recently developed algorithm (Bondy et al., 2018). For any differences in microsaccade parameters as a function of choice, we checked for any correlations with observed choice correlations.

6.3.15.3. *Effects of vergence*

We note that contrary to a stimulus without disparity noise for which small con- or divergence eye movements could result in systematically different effects on firing rate for near and far-preferring neurons (cf. Cohen & Newsome, 2008), the disparity noise in our stimulus was drawn from a wide distribution of values and designed to avoid such systematic effects. The disparity noise had a distribution of values that typically well exceeded the peak of the disparity tuning curve of the recorded units. Differences in the animals' vergence with choice therefore could have inconsistent effects on the neuronal responses to the noise stimulus depending on their tuning curves. In control analyses we examined whether any choice-related differences in vergence may have contributed to the choice-correlations we observed, in particular to the choice-correlations for the irrelevant stimulus. To examine any potential effect of small

differences in vergence across choices without making assumptions about the underlying tuning curves we directly measured for any changes in firing rate with vergence within choice. We then regressed out any vergence-dependent change in firing from the spike count for each trial and unit and re-computed choice-correlations for these corrected values.

6.3.16. *Exploring the effect of the stimulus in the opposite hemifield on firing rate*

Units were only included if they exhibited significant disparity tuning ($p<0.05$ in a one-way ANOVA). We tested for differences in mean firing rate during each 450ms stimulus presentation for each unit during a fixation task as a function of the stimulus disparity inside the receptive field and in the opposite hemifield using 2-way ANOVAs.

6.3.17. *Pair-wise interneuronal covariability*

Pairs of simultaneously recorded units were included if the units were separated by at least 100μm. Spike count correlations ("noise correlations") were computed for the average response during the stimulus presentation as the Pearson correlation coefficient between the responses of each pair to an identical stimulus. We required a minimum of 20 presentations per stimulus-condition to be included in this computation and then used the average correlation coefficient across stimuli for each pair. Tuning similarity was quantified as "signal correlations" (Bair et al., 2001) by computing the Pearson correlation coefficient between the tuning curves along the stimulus dimension used for the task (i.e. the mean response to each stimulus as a function of % signed disparity signal) for each pair of units. We quantified the relationship between noise and signal correlation by type II linear regression and significance by resampling (1000 repeats). Our results in Fig. 15f-g were similar when noise correlations were instead computed by first converting the spike counts for each stimulus condition into z-scores and then calculated as the Pearson correlation coefficient across z-scores.

6.3.18. *Statistical modeling for population analysis*

To distinguish the pattern of activation of the recorded neural populations driven by stimulus and choice – despite their correlation across trials, we used a generalized linear model (GLM) with an identity link function to predict each neuron's spike count in each trial. The GLM predicted the spike count on a given trial t based on: (1) the number frames $n_t(x)$ that each disparity x was presented in the neuron's receptive field,

$n_t^{(r)}(x)$ or $n_t^{(i)}(x)$ depending on whether the stimulus inside the receptive field was relevant or irrelevant, (2) the animal's choice on that trial $c_t^{(r)}$ or $c_t^{(i)}$, again depending on whether the stimulus inside the receptive field was relevant or irrelevant (3) whether the relevant stimulus was in the receptive field or in the opposite hemisphere (see above distinctions between *r* and *i*); (4) an estimate of the drift *d(t)* in the firing rates over the recording. An identity link function resulted in the best fits, and as a result each weight has units of firing rate modulation. More complex, nonlinear models (Whiteway & Butts, 2019) had no better model performance than the GLM for predicting the whole-trial spike count:

$$(3) \quad R(t) = \sum_x f_r(x)\, n_t^{(r)}(x) + \sum_x f_i(x)\, n_t^{(i)}(x) + w_r c_t^{(r)} + w_i c_t^{(i)} + d(t)$$

where *d*(*t*) is the model's drift term estimated using a set of tent-basis functions $[\xi_j(t)]$ (McFarland et al., 2013) that span all trials (see below) and allow a smooth set of linear model terms, $[d_j]$, as follows:

$$(4) \quad d(t) = \sum_j d_j\, \xi_j(t)$$

Thus, the predictors of the model are defined as follows (for trial *t*):

$n_t^{(r)}(x)$: the histogram of disparities on trials when the stimulus inside the receptive field was relevant, corresponding to the number of frames that the disparity *x* was presented within the receptive field of the neuron over the trial. Thus, for trials on which the relevant stimulus was presented outside the unit's receptive field, all entries of this vector are set to 0.

$n_t^{(i)}(x)$: the histogram of disparities presented on trials when the stimulus inside the receptive field was irrelevant, corresponding to the number of frames that the disparity *x* was presented outside the receptive field of the neuron over the trial. Thus, for trials on which the irrelevant stimulus was presented outside the unit's receptive field, all entries of this vector are set to 0.

$c_t^{(r)}$: the choice on trials where the stimulus inside the receptive field was relevant (near=-1, far=+1). If the stimulus inside the receptive field was irrelevant this predictor was set to 0.

$c_t^{(i)}$: the choice on trials where the stimulus inside the receptive field was irrelevant (near=-1, far=+1). If the stimulus inside the receptive field was relevant this predictor was set to 0.

$[\xi_j(t)]$: a tent basis (also known as b0-splines) that allows for a smoothly varying drift term (fittable with linear model components) to capture non-stationary aspects of the recording of each neuron not tied to any of the other predictors. There is a basis function for each anchor point *j*, with anchor points chosen with one per cycle of relevant-irrelevant stimulus blocks. The basis function is equal to 1 at the anchor point, and linearly descends to hit zero at the previous (*j*-1) and next (*j*+1) anchor points and is zero everywhere else. Thus, the corresponding model weight d_j gives the value of the offset for the model *d*(*t*) at the anchor point, and linearly interpolates at intermediate trials between the values at the anchor points.

Considering the model predictors, the model terms were constrained as follows. For a given trial, linear weights *f*(*x*) acting on the histogram of the number of video frames for which the central disc was shown at each disparity $n_t^{(.)}(x)$ yields the average number of spikes evoked by that disparity within the trial. Smoothness of the resulting tuning curve was enforced through regularization using a penalty term on the Laplacian of the weights (e.g., McFarland et al., 2013). Choice effects were fit with a single linear weight (one predictor each when the stimulus inside the receptive field was relevant or irrelevant, respectively) acting on a value corresponding to the animal's choice (-1 = near, 1 = far), and thus the model weights *w* reflected the difference in firing rate resulting from the animal's choice on that trial, again separately for trials when the stimulus was relevant w_r and irrelevant w_i. These model terms were fit simultaneously with the "drift term", which captures slow non-stationarities in the firing rate over the recording. As described above, it is fit using parameters that specify the value of the firing rate offset at each anchor point d_j, and through the use of tent-basis functions, the value of the offset linearly interpolates between "anchor points" spaced at every

period of relevant/irrelevant blocks (roughly every 94 trials; one anchor point per period).

The model parameters were fit simultaneously using gradient descent of the mean-squared error, using custom Python code. By fitting these terms all at once, the GLM could attribute the sources of the modulation to each of these factors, even though some were correlated. Such an encoding approach provides weights for stimulus and choice modulation in the same units.

6.3.19. *Analysis of stimulus- and choice-driven population activity*

The population vectors were calculated for each recording separately, from the weights of the model fits, with each dimension corresponding to a different neuron. The stimulus vector was based on the difference in weights in response to the near and far signal disparities. That is, for each neuron we computed $f_r(x=\text{near signal})$-$f_r(x=\text{far signal})$, which (across neurons) defined a vector in the neuronal population space (Fig. 4b). The population vectors for the irrelevant stimulus were computed analogously from the weights for the irrelevant stimulus. The angles, Θ, between a given pair of population vectors was calculated based on the vector dot-product:

$$(5) \quad \Theta = \cos^{-1} \frac{\boldsymbol{v}_1 \cdot \boldsymbol{v}_2}{|\boldsymbol{v}_1||\boldsymbol{v}_2|}$$

This can be visualized as the angle between lines drawn between the points in population space and the origin (e.g., Fig. 16b).

In addition to the inclusion criteria applied to the main dataset, the population vector analysis excluded units that had a majority of their response (>50%) explained by the drift term in its encoding model. Following this additional screening, only recordings with five or more valid units per area were included in the resulting measurements of population vectors. For the included units the GLM accounted for an average of 52% (V2) and 50% (V3/V3a) of the variance (five-fold cross-validation).

6.4. Results

To test these predictions, we trained two macaque monkeys to perform a disparity discrimination task on a random-dot stereogram (RDS) while ignoring another RDS (Fig. 1j). The relevant stimulus was cued block-wise, and the irrelevant stimulus was

presented in the opposite visual hemifield. It was statistically identical but independent of the relevant stimulus to ensure that it provided no information about the correct choice.

6.4.1. *The animals' behavior is spatially selective*

The animals' psychophysical behavior shows that they learned to successfully ignore the task-irrelevant stimulus (Fig.11a, c).

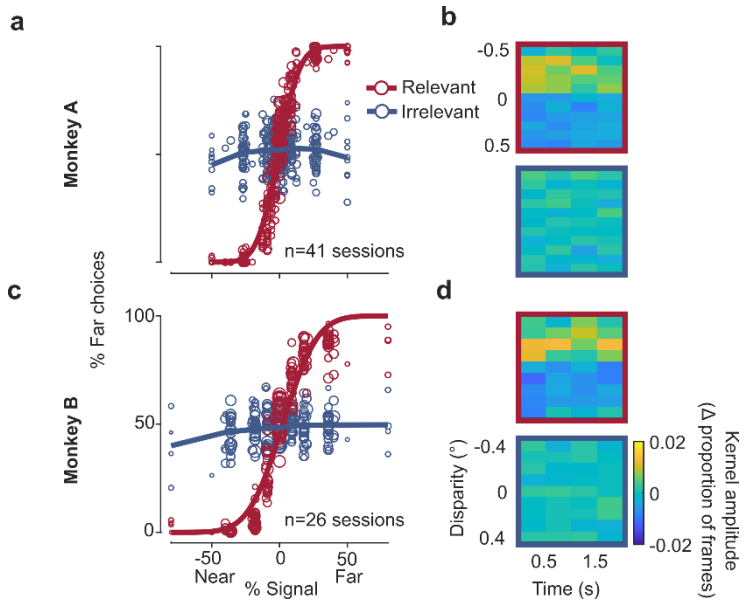

Fig. 11 | Behavior. (**a**, **c**) Psychophysical performance for each session (n=41, n=26 for monkeys A and B respectively; circles, size proportional to number of trials) for both monkeys and for the relevant (red, cumulative Gaussian fit, threshold: 12% (19%) signal for monkey A (B)) and irrelevant (blue) stimulus. **b**) Psychophysical kernels (in proportion of frames per 0.5 sec time-bin) computed for 0% signal trials (relevant, upper panel; irrelevant, lower panel) as a function of disparity and time for monkey A (trials: n=5709 relevant, n=5575 irrelevant). **d**) Same as **b**) but for monkey B (trials: n=5355 relevant, n=5203 irrelevant).

This was further verified by 'psychophysical reverse correlation' analysis, which was computed using trials restricted to the randomly rewarded no-signal trials. This analysis examines any systematic relationship, on average, between the noise in the stimulus and the animals' choices, by computing a 'psychophysical kernel' (see Methods). Non-zero values of the psychophysical kernel reveal systematic differences of the noise disparities with choice, as can be seen for the relevant stimulus (Fig. 11b, d top panels), similar to previous findings (Nienborg & Cumming, 2009). In contrast, the amplitude of the psychophysical kernel for the irrelevant stimulus was consistently around 0, confirming that the irrelevant stimulus did not systematically affect the

animals' choices. The results were similar when we selected trials for which both the relevant and irrelevant stimulus had no signal (Fig. 12).

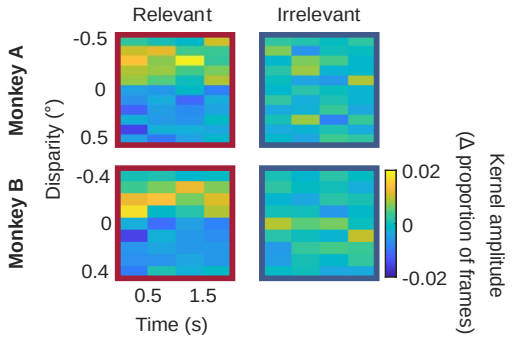

Fig. 12 | Psychophysical kernels for bilateral no-signal stimuli. For monkey A the kernels were computed for a total of n=1419 trials in 41 sessions, and for monkey B for a total of n=1328 trials in 26 sessions.

Additionally, we examined the degree to which the irrelevant stimulus or its interaction with the relevant stimulus influenced the animals' behavior using generalized linear model (GLM) analysis (see Methods). This analysis yielded weights for each covariate (the relevant stimulus, the irrelevant stimulus and their interaction). The weights for the irrelevant stimulus and for the interaction between the stimuli were close to zero (Fig. 13), which also shows that the irrelevant stimulus had minimal influence on the animals' behavior.

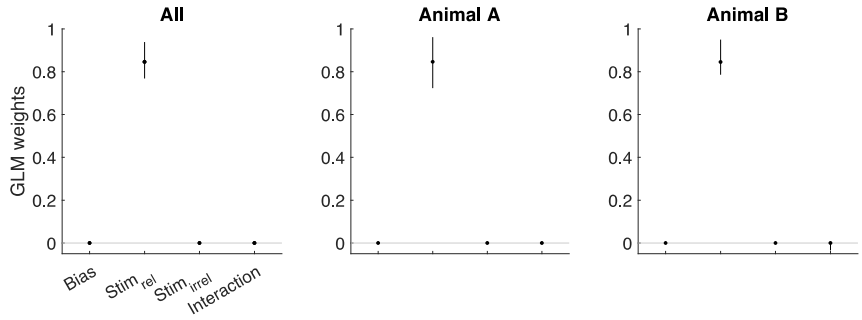

Fig. 13 | GLM analysis of behaviour. GLM weights were obtained for the animals' bias, the relevant stimulus ($Stim_{rel}$), the irrelevant stimulus ($Stim_{irrel}$) and the interaction between the relevant and irrelevant stimulus ($Stim_{rel}$*$Stim_{irrel}$). The median weights are plotted across all sessions (left panel, n=67 sessions), animal A (middle panel, n=41 sessions) and animal B (right panel, n=26 sessions), respectively. Errorbars are 95% confidence intervals around the median obtained by resampling. Note that the median weights for the irrelevant stimulus and the interaction are 0, further supporting the minimal influence of the irrelevant stimulus on the animals' choices.

6.4.2. *Visual responses are rate modulated by spatial attention*

Furthermore, as is characteristic of the modulation of visual responses by spatial attention (McAdams & Maunsell, 2000), we found a substantially smaller response when the irrelevant compared to the relevant stimulus was in the receptive field of a unit (Fig. 14a). This modulation of the neuronal response was very consistent across the populations of visual neurons in V2 and V3/V3a (Fig. 14b, V2 mean AI=0.14, n=703, p<10-30; V3/V3a mean AI=0.12, n=486, p<10-30, Wilcoxon signed rank test for significant deviation from 0).

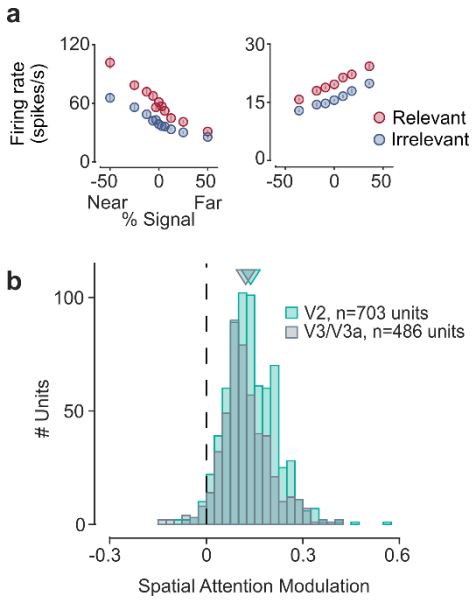

Fig. 14 | Spatial attention modulation. **a**) Tuning curves (spike rate as a function of % disparity signal; errorbars show s.e.m, n=16-111 repeats per condition) of two example units (left: V2 unit, monkey A, right: V3/V3a unit, monkey B) when the relevant (red) or irrelevant (blue) stimulus was inside the receptive field. Responses are reduced for the irrelevant stimulus as expected for modulation by spatial attention. **b**) Spatial attention index across the population for V2 (cyan, n=703 units, mean=0.14) and V3/V3a (grey, n=486 units, mean=0.12).

6.4.3. *Evidence for spatially unselective decision-related feedback*

We next examined how modulation of neural activity by decision-related feedback depended on whether the stimulus was relevant or irrelevant. The behavioral analyses satisfy a key prerequisite for this analysis, since they show that if the irrelevant stimulus had any effect on the animals' choices such effect was very small. We addressed this question using recordings in mid-level areas V2, the earliest site in the visual processing hierarchy for which systematic decision-related activity in disparity-based tasks have been observed (Nienborg & Cumming, 2006), and a subsequent processing stage, areas V3/V3a, with significant disparity selectivity (Hubel et al., 2015). We computed choice-correlations for each unit separately for trials when the task-relevant (Fig. 15a, x-axis) or task-irrelevant (Fig. 15a, y-axis) stimulus was in its receptive field. Choice-correlations quantify the degree to which a unit's firing is correlated with an animal's choice and are closely related (see Methods) to "choice

probabilities" (Federer et al., 2021) (aROC). The choice-correlations are signed. That is, positive choice-correlation values mean that a unit has a higher firing rate on trials when the animal chooses this unit's preferred disparity. Conversely, negative choice-correlations imply lower firing rates on trials when the animal chooses the unit's preferred disparity. We note that positive choice-correlations are expected for e.g. self-reinforcing feedback (Haefner et al., 2016; Wimmer et al., 2015), while negative choice-correlations would be expected, e.g. for predictive coding (Lee & Mumford, 2003). In contrast with what would be predicted in the case of selective feedback (Fig. 9a), we found that units were significantly correlated with choice even when the stimulus in their receptive field was irrelevant to the behavior. On average, the choice-correlations for the irrelevant stimulus were positive across the population both in V2 (mean=0.07, n=703, sign-rank test for significant deviation from 0: $p<10^{-12}$; monkey A: mean=0.09, n=543, $p<10^{-11}$; monkey B: mean=0.03, n=160, p=0.049) and V3/V3a (mean=0.07, n=486, $p<10^{-12}$; monkey A: mean=0.10, n=315, $p=10^{-11}$; monkey B: mean=0.02, n=171, p=0.07). (Note that decision noise during task performance would lead to worse performance and lower choice-correlations, consistent with what we observe in monkey B compared to monkey A.) Across units the choice-correlations for the relevant and irrelevant stimulus were strongly correlated both in V2 (Spearman's rank correlation, r=0.61, $p<10^{-71}$; monkey A: r=0.66, $p<10^{-68}$; monkey B: r=0.32, $p<10^{-4}$) and in V3/V3a (r=0.42, $p<10^{-21}$; monkey A: r=0.38, $p=10^{-13}$; monkey B: r=0.34, $p<10^{-5}$). This finding is incompatible with the predictions for selective feedback (Fig. 9a) or feedforward accounts (Fig. 9d) but predicted for spatially unselective task-related feedback (Fig. 9g). It also provides support for the hypothesis that feature-based attention, which is spatially global, and decision-related feedback in our task, engage a linked neural mechanism.

Conversely, the degree to which units were modulated by spatial attention was not related to their choice-correlations for either stimulus in V2, and only modestly in V3/V3a (V2: both animals: r=0.06, p=0.11; r=0.05, p=0.17 for Spearman's rank correlation between AI and the cc for the relevant and irrelevant stimulus, respectively; animal A: r=0.024, p=0.58 and r=0.018, p=0.68, respectively; animal B: r=-0.073, p=0.36 and r=0.012, p=0.88, respectively; V3/V3a: both animals: r=0.16, p=0.001 and r=0.11, p=0.01, respectively; animal A: r=0.16, p=0.01 and r=0.18, p=0.001, respectively; animal B: r=0.12, p=0.13 and r=-0.08, p=0.3, respectively).

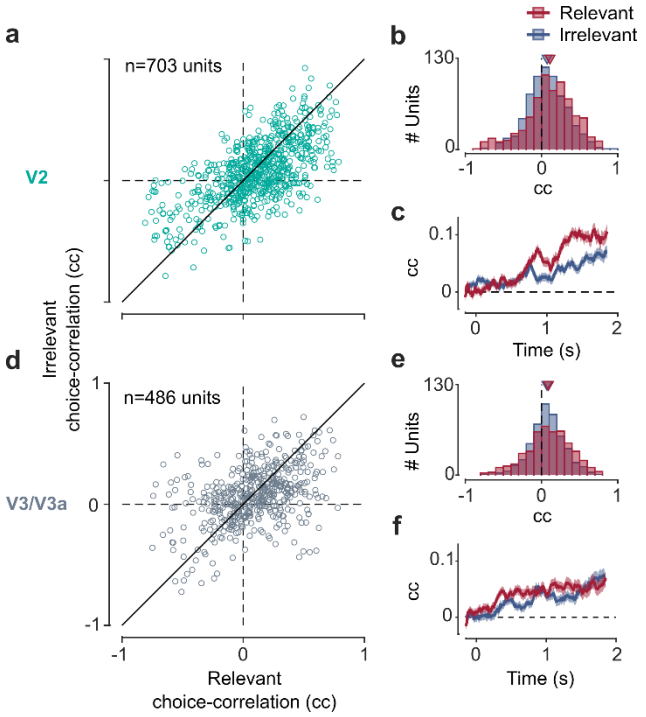

Fig. 15 | Neurons representing the task-irrelevant stimulus are correlated with choice. Choice-correlations (cc) for n=703 units in V2 when the stimulus in the receptive field was relevant (abscissa) or irrelevant (ordinate) are correlated (r=0.61, $p=10^{-71}$, two-sided Spearman's rank correlation). **b**) Choice-correlations for both the relevant (red, mean=0.11, (aROC=0.55), $p=10^{-21}$) and irrelevant (blue, mean=0.07 (aROC=0.53), $p=10^{-12}$, two-sided Wilcoxon signed rank) stimulus were significantly positive and similar although their distributions differed significantly from one another (two-sided Wilcoxon signed rank, $p=10^{-4}$). **c**) Choice-correlation (300ms wide sliding window, 10ms increment) as a function of time after stimulus onset for the relevant (red) and irrelevant (blue) stimulus. Errorbars depict s.e.m. **d**-**f**) Same as **a**-**c** but for n=486 units in V3/V3a (correlation between relevant and irrelevant choice-correlations r=0.42, $p=10^{-21}$; relevant: mean=0.09, (aROC=0.54), $p=10^{-10}$; irrelevant: mean=0.07, (aROC=0.53), $p=10^{-12}$; difference in distributions, p=0.07).

The distributions of choice-correlations for the relevant and irrelevant stimulus were overall similar although significantly different from each other in animal A in V2 (Wilcoxon signed rank, $p<10^{-3}$, p=0.08, $p<10^{-4}$ for animal A, B and both, respectively), but not in V3/V3a (p=0.26, p=0.11, p=0.07 in animal A, B and both, respectively). The difference was not explained by the stimulus selectivity of the units (V2: r=-0.007, p=0.87, r=-0.1, p=0.21; r=0.024, p=0.52; V3/V3a: r=0.019, p=0.74, r=0.13, p=0.083, r=0.06, p=0.23; for animal A, B and both, respectively, Spearman's rank correlation between the difference of the Fisher-transformed choice-correlations for the relevant and irrelevant stimulus and each unit's d-prime). Nonetheless, any such difference could reflect the feed-forward contribution to choice-correlations for the relevant stimulus, since the animals used the relevant, but not the irrelevant, stimulus for their

decisions. Our animals' behavior relied more strongly on the early part of the relevant stimulus (Fig. 11b, c). A feed-forward component, revealed as a difference between the choice-correlations for the relevant and irrelevant stimulus, should therefore be more pronounced early during the trial (cf. Nienborg & Cumming, 2009; Wimmer et al., 2015). Yet the difference in choice-correlations in our data emerged later (Fig. 15c, f). It therefore suggests that it is not entirely attributable to the feed-forward component and also reflects weakened feedback to the neurons representing the irrelevant stimulus. Since the difference emerged later during the trial although the animals' behavior relied more strongly on the stimulus early during the trial, it occurred at a time when an effect on behavior would likely be weak. Nonetheless, it is conceivable that for tasks that further increase the pressure for spatially selective processing, the decision-related feedback might be more spatially selective.

6.4.4. *Unselective modulation is not explained by the stimulus*

In control experiments and analyses we verified that the choice-correlations for the irrelevant stimulus did not result from stimulus-driven, eye-movement, or task-independent effects. First, we found no effect on a unit's firing rate of the stimulus in the opposite hemifield while the animal performed a simple fixation task that would explain the correlations with choice (see example units Fig. 17b). Indeed, the recorded units showed no systematic difference if the stimulus outside the receptive field had the preferred or null disparity (Fig. 17c). Across the population of the responses for units in V2 and V3/V3a, the proportion of units showing a modulation by the disparity outside the receptive field did not exceed chance-level (5% of units for V2, n=427, and V3/V3a, n=329, respectively, two-way Anova at 5% significance level).

6.4.5. *Unselective feedback is not explained by eye movements*

Second, our stimulus and task were designed to minimize systematic differences of eye movements with choice by disentangling saccade direction and choice (Fig. 9j, see Methods), and to minimize systematic effects of vergence on the neuronal responses. Nonetheless, we ran control analyses to show that these factors could not explain the neuronal correlations with choice. While there were no differences in microsaccade amplitude or direction with choice ($p=0.21$, $p=0.84$, respectively), microsaccades occurred with slightly higher frequency on trials preceding far choices (1.2 vs 1.4Hz, $p=10\text{-}8$, all two-sided sign rank tests). If this difference contributed to the observed

choice-correlations it would predict higher choice-correlations on sessions for which the choice-dependent difference in microsaccades was more pronounced. In contrast to this prediction, we observed a negative correlation between choice-correlation and |difference in microsaccade frequency with choice for units in V2 but not V3, V3a (131 V2: r=-0.20, p=10-7, r=-0.12, p=0.001, correlation with choice-correlation for the cued and uncued stimulus, respectively; respective values for V3/V3a: r=0.07, p=0.14, r=-0.05, p=0.28; all Spearman's rank correlation). Choice-correlations therefore tended to be higher for sessions in which microsaccades were balanced across choices. This shows that rather than contributing to choice-correlations in V2, differences in the microsaccade frequency with choice may have obscured stronger correlations with choice on some sessions.

Similar to previous studies, both animals had a tendency to converge slightly more on trials for which they made near reports (mean vergence difference for near compared to far choices across sessions: v=-0.012°, p=10-5; animal A: Δv=-0.016°, p=10-4, animal B Δv=-0.009°, p=0.029, all two-sided sign-rank tests). The resulting difference in vergence means that when the animal chose near the stimulus disparities had a tendency to be slightly shifted towards far values. The direction of this shift was therefore the opposite direction to what one would expect could result in the correlations with choice we observed. Correcting for firing-rate differences induced by vergence (see Methods) resulted in very similar but on average slightly increased values for choice-correlations compared to the uncorrected values (corrected mean choice-correlation and correlation with the uncorrected values in V2: ccrelevant=0.115, r=0.96, p=10-20, ccirrelevant= 0.088, r=0.95, p=10-20; in V3: ccrelevant=0.075, r=0.94, p=10-20, ccirrelevant= 0.072, r=0.91, p=10-20), as expected from the direction of the vergence difference. It shows that the choice-correlations we measured are not accounted for by vergence eye movements but rather that due to the animals' vergence we may have slightly underestimated the neuronal correlations with choice.

6.4.6. *Unselective feedback is not explained by differences between stimuli*

Although we corrected the choice-correlations for systematic stimulus differences with choice (see Methods), we verified that they were similar when we removed any differences in the randomly generated stimuli between trials (Fig. 16). Although this analysis used only half of the trials, rendering the data noisier, we nonetheless

observed similar results for for the frozen noise stimuli as those for the corrected choice-correlations. The values of the choice-correlations corrected for stimulus-induced variations were highly correlated with those obtained for frozen noise (V2: $r_{relevant}$=0.46, $p<10^{-13}$, n=239; $r_{irrelevant}$=0.29, $p<10^{-4}$, n=209; V3: $r_{relevant}$=0.71, $p<10^{-20}$, n=143; $r_{irrelevant}$=0.52, $p<10^{-10}$, n=139, Spearman's rank correlations). While the correlation between the choice-correlations for the neurons representing the relevant and irrelevant stimulus, respectively, was reduced to a trend in this subset of the data in V2 (n=209, r=0.10, p=0.14, Spearman's rank correlation), it remained highly significant across the V3/V3a population (r=0.28, p<0.001), similar to the corrected values obtained for the entire population.

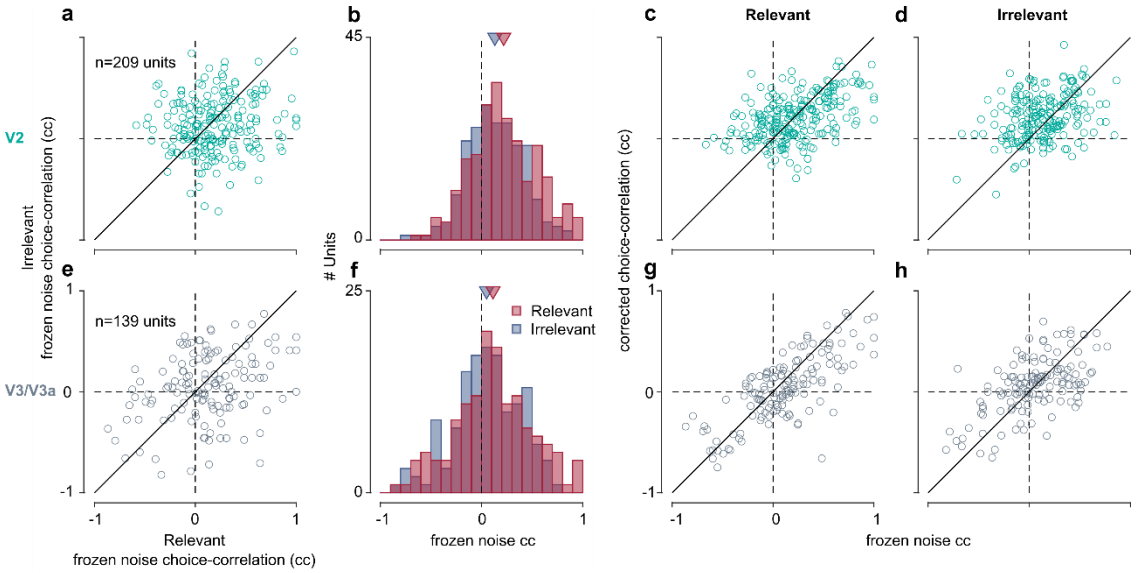

Fig. 16 | Choice-correlations are similar for identical stimuli ("frozen noise"). **a**) Choice-correlations (cc) were calculated for trials with a fixed sequence of disparity frames ("frozen noise") and compared for the relevant (abscissa) or irrelevant (ordinate) stimulus. While the correlation was weak for V2 (n=209, r=0.10, p=0.14, Spearman's rank correlation), it was highly significant across the V3/V3a population (**e**, r=0.28, p<0.001), similar to the corrected values obtained for the entire population. **b**) Choice-correlations for frozen noise were significantly positive for both the relevant (mean=0.19, $p<10^{-13}$) and irrelevant stimulus (mean=0.13, $p<10^{-9}$, both Wilcoxon sign-rank test) and comparable, although their medians differed significantly from one another (Wilcoxon rank sum, p=0.03). **c**) Choice-correlations calculated for frozen noise were similar to those obtained for the white noise corrected for variations in the stimulus for the relevant stimulus (r=0.46, $p<10^{-13}$, n=239) and (**d**) irrelevant stimulus (r=0.29, $p<10^{-4}$, n=209). **e-h**) Same as **a-d** but for V3/V3a (n=139; relevant: mean=0.12, p<0.001; irrelevant: mean=0.05, p=0.04; difference in distributions, p=0.13; correlation of choice-correlations for frozen stimuli with those corrected for stimulus variations for the relevant stimulus: r=0.71, $p<10^{-20}$, n=143 and irrelevant stimulus: r=0.52, $p=10^{-10}$, n=139).

6.4.7. *Unselective modulation is not explained by task-independent noise correlations*

Third, our findings could not be accounted for by stimulus-independent correlated variability (“noise correlations”, Cohen & Kohn, 2011; Zohary et al., 1994) across hemispheres. Such an explanation (e.g. Haefner et al., 2013; Nienborg et al., 2012; Shadlen et al., 1996) would stem from previous observations that pairs of neurons with similar stimulus tuning tend to have higher noise correlations than those with dissimilar tuning (Cohen & Maunsell, 2009; Kohn & Smith, 2005; Zohary et al., 1994). Indeed, we found such a relationship in our data when the animals were fixating and not engaged in the task: within hemisphere the correlated variability increased with signal correlation (V2: r=0.08, [0.07 0.1], n=3847, V3/V3a: r=0.11, [0.09 0.13], n=1519, type II regression, 95% CI; Fig. 17f, left column, blue). However, across hemispheres we observed no systematic relationship between correlated variability and signal correlation in V2 (r=0.01, [-0.001 0.02], n=1726) or V3/V3a (r=-0.01, [-0.02 0.002], n=660; Fig. 17f, left column, orange). The correlated variability during fixation therefore lacked the structure that would be required to account for choice-correlations in a feedforward way, suggesting that the choice-correlations for the task-irrelevant stimulus result from decision-related feedback.

In fact, the presence of choice-correlations for the task-irrelevant stimulus implies (Wasmuht et al., 2019) that neurons in both hemispheres that share tuning preferences should also receive more similar decision-related feedback compared to inter-hemispheric pairs with different tuning. This would predict that, during task performance, there should be increased noise correlations between neurons in different hemispheres that have similar tuning (i.e. signal correlation). Our results showed exactly this (Fig. 17g; V2: r=0.11, [0.02 0.22], n=711 within hemisphere, r=0.12, [0.07 0.17], n=450 across hemispheres; V3/V3a: r=0.25, [0.14 0.34], n=461 within hemisphere and r=0.26, [0.16 0.37], n=136 across hemispheres). The fact that the structure of the correlated variability changes when the animal engages in the task further supports the finding that this structure depends on task-related feedback (Bondy et al., 2018).

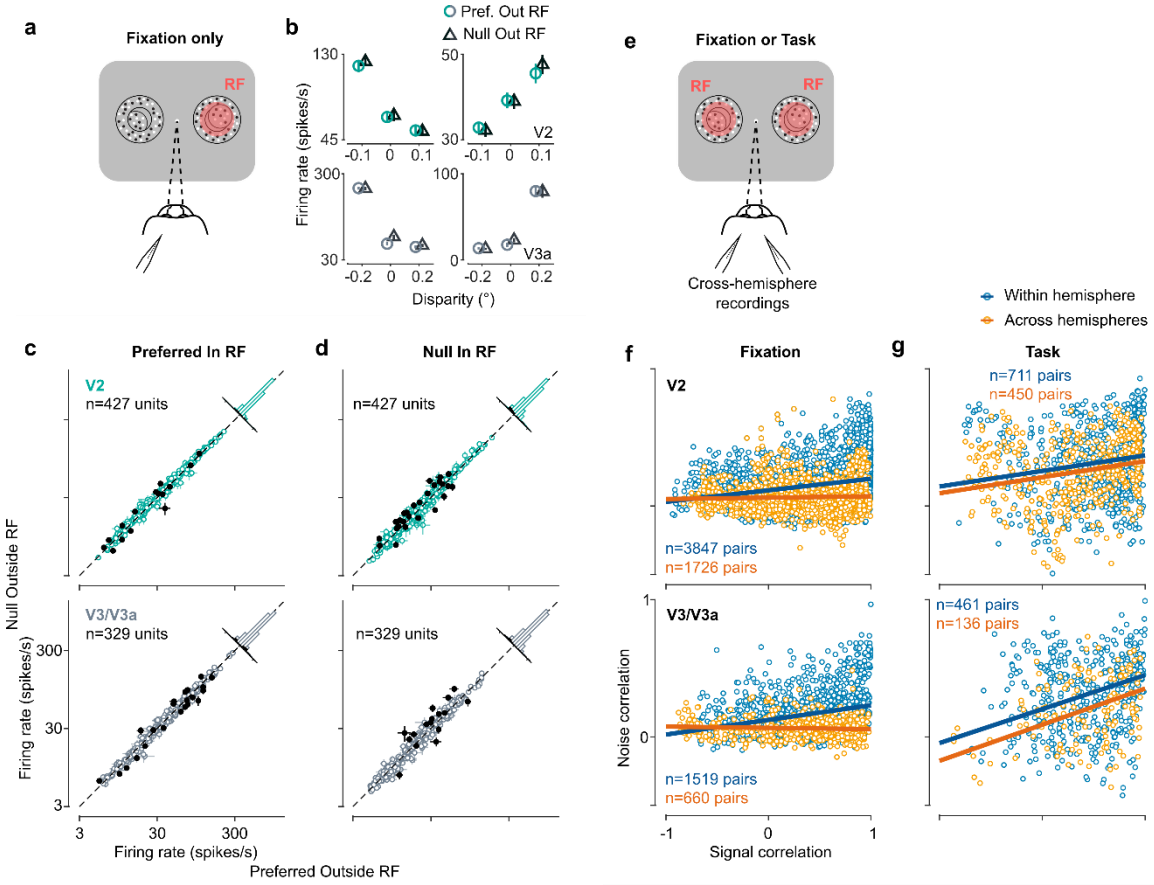

Fig. 17 | Choice-correlations for the irrelevant stimulus do not result from the stimulus outside the receptive field, nor from noise correlations in the absence of a task. a) Fixation task with one stimulus presented inside the receptive field, and one stimulus in the opposite hemifield. **b**) Disparity tuning curves of four example units (V2, top row; V3/V3a, bottom row) as a function of the disparity of the stimulus inside the receptive field, plotted separately (labels) for the disparity of the stimulus in the opposite hemifield. Data points are horizontally offset for visibility. Errorbars show s.e.m. n=10-40 repeats per condition) **c**) Firing rates to the preferred disparity (top: V2, bottom: V3/V3a) inside the receptive field as a function of the disparity of the stimulus in the opposite hemifield. Histograms show firing rate ratios. Filled data points depict units whose firing rates significantly deviate from unity ($p<0.05$, two-sample t-test). Error bars show s.e.m. **d**) as **c**) but for the null disparity inside the receptive field. **e**) Bi-hemisperic recordings during fixation or task performance. **f**) Noise correlations between pairs of units within (blue) and across (orange) hemispheres are plotted as a function of their signal correlation, during fixation and **g**) during performance of the task for V2 (top) and V3/V3a (bottom). Note the change in the regression slope within versus across hemispheres during fixation but not during task performance.

6.4.8. *Analysis of population responses*

To better understand how the structure of the decision-related variability related to the stimulus representation across the population (cf. also Ni et al., 2018), we used a

generalized linear model (GLM) to separately estimate the effect of stimulus tuning and decision-related modulation on neural activity. The model was necessary because the choice and stimulus were often correlated across trials, and the GLM assigns weights according to the best explanation for each effect, which are statistically separable from each other despite their correlation. This model predicted the response of a given neuron (its spike count) on each trial t using the distribution of disparities per trial, the animal's choice and a drift term (Fig. 18a, see Methods).

The main parameters of the model used to predict firing rate are the disparity tuning curves for the relevant and irrelevant stimulus, $f_r(x)$ and $f_i(x)$ and the choice weights in each condition weights w_r and w_i. Each tuning curve $f(x)$ is comprised of weights (one weight per disparity x), which operate on the histogram of disparities presented over a given trial $n_t(x)$ (Fig. 16a, left), and the choice weights w operate on the choice (Fig. 18a, middle). We also fit a drift term $d(t)$ to segregate any non-stationary effects in the recording, which was constrained to be slowly varying across a session, i.e. one parameter for each cycle of relevant and irrelevant blocks (see Methods). For the tuning curves and choice weights, separate weights were fit to trials when the stimulus inside a unit's receptive field was relevant versus irrelevant, while the same time-varying drift term was applied to all trials.

The best model predictions were accomplished without additional nonlinear mappings (e.g. a spiking nonlinearity link function in the GLM, see Methods). Therefore, the model terms that reflect how much each unit was driven by the stimulus and choice are in units of firing rate, and thus can be directly compared across units. This allowed us to examine the degree to which the changes in activity with choice or stimulus were aligned at the level of the population. Specifically, each pattern was represented by a population vector in an M-dimensional space, where M is the number of simultaneously recorded units in a given experimental session, and each axis corresponds to the response of one neuron (Fig. 18b). Thus, similar patterns of activity correspond to similar "directions" in population space (relative to the origin) and there would be a correspondingly small angle between them.

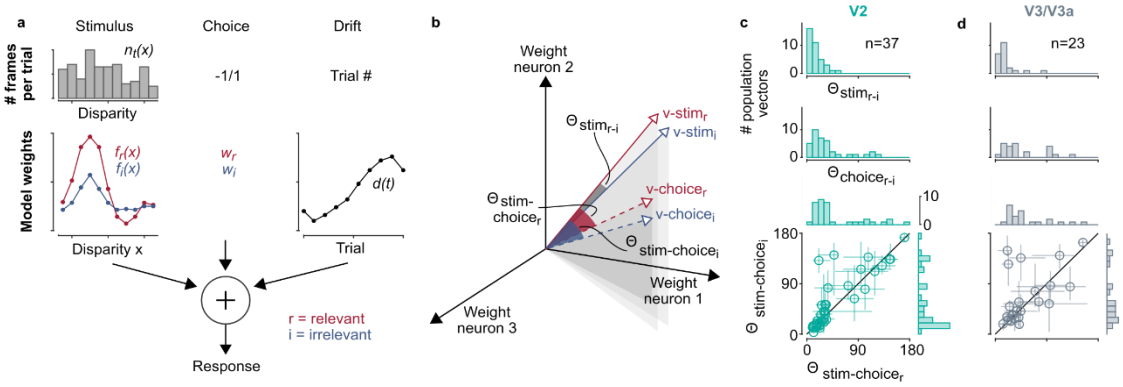

Fig. 18 | Stimulus and choice alignment of the population response in visual cortex. **a**) The encoding model predicts the spike count in a trial from the stimulus-driven response based on its disparity tuning curve and the disparities presented in that trial (left), the choice during that trial (middle), and a drift term to fit any non-stationarity in the recording over time (right). **b**) The stimulus- or choice-driven patterns of neural activity across the recorded neurons can be represented as population vector, with each dimension corresponding to a given neuron's stimulus- or choice-driven weight estimated from the encoding model. The patterns of activity can then be compared as an angle between the population vectors. **c**, **d**) Results for populations in V2 and V3/V3a. Top: The population vectors for the relevant and irrelevant stimulus are well aligned across sessions. Middle: the population vectors for the choices when the stimulus was relevant or irrelevant are broadly aligned. Bottom: The angles between the population vectors for the relevant stimulus (v-stim$_r$) and choice (v-choice$_r$ or v-choice$_i$) are correlated (r=0.86, p=10^{-10}, n=37 sessions for V2; r=0.51, p=0.01, n=23 sessions for V3/V3a, two-sided Spearman's rank correlation; errorbars are 95% confidence intervals around the median angle, by resampling).

6.4.9. *Stimulus-choice (mis)alignments are similar across relevance*

We considered four population response vectors for the changes between the near signal disparity and far signal disparity in the stimulus (see Methods for details) when it was relevant (Fig. 18b, **v-stim$_r$**), irrelevant (**v-stim$_i$**), and for the changes in the population response with choice when a relevant (**v-choice$_r$**), or irrelevant (**v-choice$_i$**), stimulus was inside the receptive field of the population. We then computed the angles between these population vectors as a measure of the extent to which the population responses were aligned between the conditions. Across sessions the relevant and irrelevant stimulus vectors ($\Theta_{stim\ r\text{-}i}$) were well aligned both in V2 (Fig. 18c, top) and V3/V3a (Fig. 18d, top). It is consistent with previously described gain changes as a function of spatial attention (McAdams & Maunsell, 1999; Reynolds & Heeger, 2009) and suggests a roughly uniform change in gain with spatial attention across the

population. But the choice vectors were less well aligned with that of the relevant stimulus ($\Theta_{stim\ r\text{-}choice\ r}$, $\Theta_{stim\ r\text{-}choice\ i}$) in a number of sessions, consistent with recent findings for population recordings in area MT (Zhao et al., 2020).

Interestingly, these misalignments between the choice vectors and the relevant stimulus vector were very consistent between the relevant and irrelevant choice vectors (Fig. 18c, d, bottom: r=0.86, $p<10^{-10}$, n=37 for V2; r=0.51, p=0.01, n=23 for V3/V3a). Moreover, the choice vectors ($\Theta_{choice\ r\text{-}i}$) were broadly aligned (Fig. 18c, d, middle). This suggests that the decision-related signal affected the population in visual cortex in a similar way, whether this population represented a relevant or irrelevant stimulus, and that this decision-related signal can be misaligned with the stimulus-representation. If sensory and non-sensory signals are multiplexed at the level of the sensory population, as previously suggested (e.g. Zhao et al., 2020), such consistency in the representation of the non-sensory decision signals may facilitate their use by downstream processing.

Additionally, the angles between choice vectors ($\Theta_{choice\ r\text{-}i}$) were not correlated with the angles between the relevant versus irrelevant stimulus vectors ($\Theta_{stim\ r\text{-}i}$) (r=0.15, p=0.37, n=37 for V2; r=-0.35, p=0.11, n=23 for V3/V3a), consistent with the weak (V3/V3a), or absent (V2) correlation between the modulation by spatial attention and choice-correlation. This also suggests that the modulation by spatial attention or choice is not prominently coupled in this task.

6.5. Discussion

In summary, we observed substantial choice-correlations for neurons representing a task-irrelevant, ignored stimulus, which could not be explained by task-independent co-variates or feed-forward sensory noise. Rather, these choice-correlations require feedback interactions that are roughly similar whether or not the stimulus inside a neuron's receptive field is relevant. From the perspective of the decision-process in this task this is remarkable. Our task was designed to eliminate uncertainty as to which stimulus was task-relevant and analyzing the animals' behavior verified their negligible use of the irrelevant stimulus. Neurons representing this irrelevant sensory information were nearly as correlated with choice as were neurons representing the sensory information that the animals measurably relied on. These findings appear to call into

question previously observed systematic links – even if they reflect feedback interactions- between sensory neurons with choice-correlations, and the perceptual decision-process (e.g. Britten et al., 1996; Nienborg & Cumming, 2007; Uka & DeAngelis, 2003; Yu & Gu, 2018). Conversely, the choice-correlations for neurons are expected for a mechanism engaging feature-based attention, as previously hypothesized. Our findings here therefore provide support for the hypothesis that the decision-related feedback is linked to the spatially global mechanism engaged in feature-based attention (Bondy et al., 2018; Cohen & Newsome, 2008; Krug, 2004).

However, while a spatially unselective mechanism is beneficial in search or detection tasks that target feature-based attention mechanisms and typically contain spatial uncertainty (Bichot et al., 2005; Cohen & Maunsell, 2011), the task used here involved no uncertainty about which location was relevant. Indeed, the measured behavior was highly spatially selective. Any lack of spatial selectivity of the decision-related feedback observed here is therefore not attributable to the demands of the task.

Our findings here extend beyond previous reports (e.g. Li et al., 2008; Mirabella et al., 2007; Steinmetz et al., 2019) of unselective task- or decision-related feedback. First, we verified behaviorally that the irrelevant information was ignored by the animals. Second, our task was designed to uncouple the decision-formation from the motor-plan to report the decision.

In this study we explored the selectivity of the feedback once the animals were fully trained on the task. Our results therefore leave open the possibility that the feedback was spatially selective during earlier phases of the training, e.g. to support the animals' learning of the task.

A lack of spatial selectivity of the decision-related feedback could have implications for downstream processing (Haefner et al., 2016; Haimerl et al., 2019; Wimmer et al., 2015). It suggests a common mechanism across tasks that is independent of the spatial selectivity that those tasks demand. It also challenges theoretical accounts for the computational role of feedback that require selectivity. The lack of selectivity may result from a biological constraint. Assume that the selectivity of the feedback could be increased to target an arbitrary number of stimulus dimensions, and this selectivity is mediated by selective wiring. This would require the number of

connections to grow exponentially with each additional stimulus or task dimension. Restricting the selectivity of the modulation by feedback, as observed here, reduces the wiring required for the feedback, and may also facilitate generalization across tasks (Ruff et al., 2018). Such biological constraints may be key to solving the longstanding puzzle of the computational role of feedback (van Bergen & Kriegeskorte, 2020) and its implementation.

7. General Discussion

7.1. Summary

The studies included in this thesis were devised to contribute to a long line of research into the role of visual neurons in perception. Our main aim was to characterise feedback in perceptual decision-making, by investigating the selectivity of the feedback that gives rise to decision-related activity. To do so we focussed on disparity perception and disparity-selective neurons in mid-level visual areas.

We first investigated whether neurons in area V2 were causally involved in disparity perception, by applying microstimulation whilst two macaque monkeys performed a disparity discrimination task (chapter 5). We found that microstimulation led to a bias in behaviour, such that on stimulation trials animals were more likely to make a decision towards the preferred disparity of the stimulated site. The size of this bias was affected by the tuning strength of the stimulated sites: stronger selectivity led to larger effects of microstimulation on behaviour. However the relationship with reward size was unclear as it differed between the two animals, and may reflect a difference in strategies between the two animals. These results suggest that disparity-selective V2 neurons are causally involved in such a disparity discrimination task, and that the more selective the neurons, the greater their contribution to the decision.

Next we aimed to probe the selectivity of feedback during perceptual decision-making using a novel task combining disparity discrimination with a spatial attention component (chapter 6). We found that although the animals' behaviour was spatially selective, only using the task-relevant stimulus for the decision, the neural activity was not. Specifically, we found decision-related activity for neurons representing the ignored, task-irrelevant stimulus. Using a set of control experiments and analyses, we showed that this could not be explained by any effect of the stimulus in the opposite hemifield, nor by noise correlations across hemispheres in the absence of the task. Crucially, this result would not be expected if feedback in this task selectively targeted neurons solely based on their task-relevance. Our finding thus rather points to a spatially global feedback mechanism similar to feature attention. Given that such spatially unselective feedback would not be required by the demands of our task, its

persistence suggests that the brain may employ a general feedback mechanism across different types of task, regardless of their specific spatial demands.

In the scope of the second project described in chapter 6, we made several additional findings of note. First, the structure of noise correlations across hemispheres changed depending on the presence or absence of the task. When the animals were not performing the task and simply fixating, noise correlations were higher between similarly tuned neurons within the same hemisphere, but this relationship did not hold across hemispheres. However, when performing the task, this relationship changed, such that both within and across hemispheres noise correlations were positively correlated with the degree of tuning similarity. This precise structure would be predicted given the link between noise correlations and decision-related activity, and our finding of the latter in both hemispheres during the task. It also provides further support for a feedback origin of noise correlations. Furthermore, through modelling of the population responses, we also found that stimulus- and decision-related activity were somewhat misaligned in population space. Importantly, this misalignment did not significantly differ depending on whether the stimulus inside the receptive field was task-relevant or task-irrelevant. This result suggests that feedback giving rise to decision-related activity affects visual neurons in a similar way, regardless of their contribution to the decision.

Although the results presented here focus on disparity perception in V2 and V3/V3a, we expect our findings to extend to other features, such as motion and orientation, and the neurons selective for them. Our work therefore provides broad insight into feedback mechanisms involved in perceptual decision-making, and invites future projects to probe the role and extent of this feedback, within this and other datasets.

7.2. Evidence for a general feature-selective feedback mechanism

Over the last 10 years, it has been suggested that the feature-selective feedback modulations of visual neurons that give rise to decision-related activity may engage the same mechanism as feature attention (e.g. Bondy et al., 2018; Cohen & Newsome, 2008; Krug, 2004; Nienborg & Cumming, 2009). Since feature-selective attention exhibits a spatially global modulation of neurons tuned for the attended feature (Bichot et al., 2005; Treue & Trujillo, 1999), it followed that, if decision-related activity also

arises from a similar mechanism, it may also be found for neurons across the visual field. This is indeed what we discovered, with prominent decision-related activity in V2 and V3/V3a neurons representing a task-irrelevant stimulus in the opposite hemifield to the task-relevant one. Our results thus provided compelling evidence of a link between feedback in a perceptual decision-making task and feature attention. Further, the critical difference between our task and those typically used for feature-selective attention is that here, animals always knew the spatial location of the task-relevant stimulus, which was clearly reflected in their behaviour. This suggests that this feature selective feedback mechanism acts in a similar manner across different types of task, regardless of the spatial specificity the task demands.

7.3. The limits of feedback selectivity

Our result also provides a substantial challenge to theoretical accounts of feedback which assume feedback selectively targets task-relevant neurons (Haefner et al., 2016; Haimerl et al., 2019). Specifically, it suggests that task-relevant feedback may be limited in the degree of its selectivity, potentially only able to select target neurons according to a small number of feature dimensions. In our case, task-relevant feedback was selective along the disparity feature dimension but not along the spatial dimension, acting seemingly independent of the modulation by spatial attention. Although at first glance this may appear to be a disadvantage for perceptual flexibility, it could to the contrary prove advantageous for two reasons. First, limiting the anatomical connections along fewer feature axes would restrict the biological cost of projecting to infinite combinations of neurons depending on the task demands. Second, fewer mechanisms that can be utilised across different tasks with the same basic building blocks could facilitate generalisation, cutting the time or energy required to learn a new task.

One matter left unclear by our results is how many areas are directly targeted by the feedback which gives rise to decision-related activity. Here we observed decision-related activity in a disparity discrimination task in areas V2 and V3/V3a. However, similar observations have also been made in area MT (Uka & DeAngelis, 2004). Similarly widespread decision-related activity can also be seen for features like heading, where the presence of this activity does not relate to whether or not an area is causally involved in the task (Yu & Gu, 2018). One possibility is that feature-selective

feedback is also limited in its areal specificity, projecting to all areas that show selectivity for the relevant or attended feature (Maunsell & Treue, 2006), regardless of their causal involvement in the task. If this were the case, this could lead to increases in decision-related activity in each subsequent area in the visual processing hierarchy, as they would not only receive the feedback themselves, but additionally inherit the feedback modulations of upstream areas in the feedforward sweep. This is supported by evidence that decision-related activity tends to increase in magnitude in downstream sensory areas (Yu & Gu, 2018; Crapse & Basso, 2015), regardless of whether the areas in question are involved in perception of the relevant feature in question (Yu & Gu, 2018). Although in our results, the choice probabilities in V3/V3a did not appear to be greater in magnitude than those in V2, further investigation is needed to provide a conclusive answer to this question. For this purpose, additional data collected during this work and simultaneously recorded in areas V2 and V3/V3a, as well as some in area V4, both within and across hemispheres, could provide a useful starting point for future study.

There is however evidence for at least one factor which may determine whether an area is targeted by this feedback: whether or not there is topographic organisation for the task-relevant feature (Nienborg & Cumming, 2014). This principle is well illustrated in area V1, which, in spite of containing individual disparity-selective neurons, does not exhibit decision-related activity in a disparity discrimination task (Nienborg & Cumming, 2006). Conversely, V1 neurons do exhibit decision-related activity in an orientation discrimination task (Nienborg & Cumming, 2014). Whilst orientation-selective neurons are organised topographically in V1 (Hubel & Wiesel, 1968), disparity-selective neurons are not (Nienborg & Cumming, 2006), and it could be this key difference which determines whether decision-related activity can be measured. The implication for feature-selective feedback is then that it may be limited by the feature-selective structure of the areas it targets, perhaps organised in such a way as to send a limited number of projections to clusters of neurons with a preference for the same feature.

7.4. The origin of feedback in perceptual decision-making

In addition to the question of how broadly this feedback acts, our work also leaves open the question of where it originates. Areas in posterior parietal cortex, such as LIP,

have historically been thought to play a role in evidence accumulation and decision formation (see review by Huk et al., 2017), whilst prefrontal areas are thought to be involved in the coordination of feature-selective attention (Bichot et al., 2019, 2015). By taking advantage of recent methodological advances which allow recordings across many areas simultaneously (e.g. Dotson et al., 2018; Siegel et al., 2015; Steinmetz et al., 2018), future investigations could hope to illuminate not only the source of this feedback, but aid our understanding of the entire network involved in making perceptual decisions.

7.5. Revisiting the role of feedback

The precise role of feedback in perception also remains unclear. Given that we observed evidence of feedback targeting neurons that were not relevant for the task, as well as those that were, our results appear to be incompatible with accounts that require feedback to specifically target task-relevant neurons only, for example those where the feedback dictates which neurons are read out at a later stage (Haimerl et al., 2019). However, our findings do not rule out the possibility that there may be a selective downstream readout of these unselective feedback modulations.

Our finding that stimulus and choice information were not entirely aligned in population activity (chapter 6) suggests that decision-related information can be read-out at least somewhat independently of the sensory driven activity, or ‘multiplexed’ (Zhao et al., 2020). This could indicate that rather than feedforward and feedback contributions being directly integrated in sensory neurons in a manner which contributes to the resulting choice (Haefner et al., 2016; Wimmer et al., 2015), choice information is processed in part independently and is in some way informative for downstream areas. Whether such a representation would be continually updated as the choice is made, or rather reflects a post-decision signal (Goris et al., 2017) is also unclear, but investigating these questions could provide further constraints on future models seeking to explore the precise computational role of feedback.

7.6. Reassessing decision-related activity

Decision-related activity in visual neurons has long been taken to imply a relationship between the visual neurons in which it is found and the decision an animal makes (Britten et al., 1996; Shadlen et al., 1996). Although more recent evidence has called

this into question (Goris et al., 2017; Yu & Gu, 2018), even some of the most recent theoretical accounts of this activity have sought to reconcile possible feedforward and feedback contributions (Haefner et al., 2016; Wimmer et al., 2015), the latter of which selectively target task-relevant neurons, implying that the presence of decision-related activity may still reflect some role for those neurons in the task. Here, not only did we find decision-related activity for neurons representing an ignored, task-irrelevant stimulus, but also that this was similar in magnitude to the decision-related activity found for neurons representing the task-relevant stimulus. These findings therefore demonstrate that decision-related activity can be predominantly explained by feedback, rather than the feedback component reflecting only a small part of this activity.

Our findings therefore also serve to reaffirm that decision-related activity cannot be taken as a proxy for the causal involvement of a particular area, echoing previous findings (Yu & Gu, 2018). Given that we observed decision-related activity for neurons representing task-irrelevant stimuli that were not used to inform the behaviour, it is clear that the presence of this activity in no way confers that those neurons are being used in the decision. This highlights the continued importance of using causal manipulations, such as microstimulation (chapter 5), to establish causal links between particular neural populations and perception in specific tasks. With recent methodological advances, such as optogenetic stimulation (Galvan et al., 2017), we can expect increasingly sophisticated spatial and temporal targeting of neuronal populations than ever before, aiding our understanding of the role of specific circuits in sensory decisions.

7.7. Outlook

As discussed above, our results leave several questions open for investigation, including the areal specificity of the feedback we measured, its origin, and the specific role it may play in perception. The data collected over the course of this thesis will no doubt continue to provide important insights into some of these questions. For example, recordings made across multiple areas simultaneously could help illuminate how decision-related activity progresses through the visual processing hierarchy, by tracking communication between areas and the timing of these modulations. Furthermore, given that our recordings were made using linear multi-electrode probes

inserted orthogonally to the cortical surface, these projects will have access to the laminar profile of decision-related activity, allowing further delineation of the feedback and feedforward contributions based on the layers in which it is found (Briggs, 2020; Rockland & Pandya, 1979).

In addition, a number of related questions beyond the scope of this work could also be addressed. For example, although our present analysis focussed on the spiking activity of neurons, analysing the local field potentials (LFP) could provide additional insights into the frequency-specific profiles of both feedforward and feedback mechanisms underlying perceptual decision-making (e.g. Fries et al., 2008; Krishna et al., 2020; Pesaran et al., 2018; Spyropoulos et al., 2018; Wilming et al., 2020). Additionally, whilst the neural data itself will provide the basis for future research, simultaneously recorded eye movements and video capture of the entire animal have already become the basis of ongoing projects. These concern, for example, the development of microsaccade profiles during the learning of a spatial attention task (Bellet et al., 2019, conference proceedings), and the question of whether uninstructed body movements affect sensory processing in the primate (Talluri et al., 2022), as they do in the rodent (Musall et al., 2019).

This serves to highlight the richness of the data I collected here. Although collected with specific questions in mind, addressed by this thesis, such invasive recordings can provide numerous insights into a wide range of phenomena far beyond those covered by an individual project.

7.8. Concluding remarks

In this thesis we investigated the role of neurons in mid-level visual areas in perception, with a specific focus on disparity. We not only demonstrated a causal role for V2 neurons in disparity perception, but also showed that decision-related activity in this area, and in V3/V3a, could be predominantly explained by feedback that targeted disparity-selective neurons regardless of their task-relevance. Our results go beyond previous evidence demonstrating that feedback accounts for at least some of the variability of visual neuronal responses. First, our findings suggest that feedback can explain more of this variability than necessarily expected. Secondly, they show that it also acts far more broadly than predicted. This illustrates how little we can rely on correlational measures to make conclusions about the role of particular visual neurons in perception, and highlights the continued importance of causal manipulations for such questions. Critically, these results provide valuable insights that can be utilised for our understanding of the specific role of feedback in perception. By characterising the limits of the selectivity of this feedback, our findings provide constraints for future theoretical models, and may provide clues as to how feedback modulations of visual neurons may or may not be used to inform perception.

8. References

Adams, D. L., & Zeki, S. (2001). Functional Organization of Macaque V3 for Stereoscopic Depth. *Journal of Neurophysiology*, *86*(5), 2195–2203. https://doi.org/10.1152/jn.2001.86.5.2195

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, *52*(6), 1106–1130. https://doi.org/10.1152/jn.1984.52.6.1106

Albright, T. D., Desimone, R., & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology*, *51*(1), 16–31. https://doi.org/10.1152/jn.1984.51.1.16

Anzai, A., Chowdhury, S. A., & DeAngelis, G. C. (2011). Coding of Stereoscopic Depth Information in Visual Areas V3 and V3A. *Journal of Neuroscience*, *31*(28), 10270–10282. https://doi.org/10.1523/JNEUROSCI.5956-10.2011

Bair, W., Zohary, E., & Newsome, W. T. (2001). Correlated Firing in Macaque Visual Area MT: Time Scales and Relationship to Behavior. *The Journal of Neuroscience*, *21*(5), 1676–1697. https://doi.org/10.1523/JNEUROSCI.21-05-01676.2001

Bellet, M., Seillier, L., Quinn, K., Kawaguchi, K., Bellet, J., Hafed, Z., & Nienborg, H. (2019). Microsaccade Patterns Evolve During Learning of a Covert Spatial Attention Task. *2019 Conference on Cognitive Computational Neuroscience*. 2019 Conference on Cognitive Computational Neuroscience, Berlin, Germany. https://doi.org/10.32470/CCN.2019.1331-0

Bichot, N. P., Heard, M. T., DeGennaro, E. M., & Desimone, R. (2015). A Source for Feature-Based Attention in the Prefrontal Cortex. *Neuron*, *88*(4), 832–844. https://doi.org/10.1016/j.neuron.2015.10.001

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and Serial Neural Mechanisms for Visual Search in Macaque Area V4. *Science*, *308*(5721), 529–534. https://doi.org/10.1126/science.1109676

Bichot, N. P., Xu, R., Ghadooshahy, A., Williams, M. L., & Desimone, R. (2019). The role of prefrontal cortex in the control of feature attention in area V4. *Nature Communications*, *10*(1), 5727. https://doi.org/10.1038/s41467-019-13761-7

Bijanzadeh, M., Nurminen, L., Merlin, S., Clark, A. M., & Angelucci, A. (2018). Distinct Laminar Processing of Local and Global Context in Primate Primary Visual Cortex. *Neuron*, *100*(1), 259-274.e4. https://doi.org/10.1016/j.neuron.2018.08.020

Bondy, A. G., Haefner, R. M., & Cumming, B. G. (2018). Feedback determines the structure of correlated variability in primary visual cortex. *Nature Neuroscience*, *21*(4), 598–606. https://doi.org/10.1038/s41593-018-0089-1

Briggs, F. (2020). Role of Feedback Connections in Central Visual Processing. *Annual Review of Vision Science*, *6*(1), 313–334. https://doi.org/10.1146/annurev-vision-121219-081716

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience*, *13*(1), 87–100. https://doi.org/10.1017/S095252380000715X

Britten, K. H., & van Wezel, R. J. A. (1998). Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nature Neuroscience*, *1*(1), 59–63. https://doi.org/10.1038/259

Britten, K., Shadlen, M., Newsome, W., & Movshon, J. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience*, *12*(12), 4745–4765. https://doi.org/10.1523/JNEUROSCI.12-12-04745.1992

Burkhalter, A., & Van Essen, D. (1986). Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *The Journal of Neuroscience*, *6*(8), 2327–2351. https://doi.org/10.1523/JNEUROSCI.06-08-02327.1986

Celebrini, S., & Newsome, W. (1994). Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. *The Journal of Neuroscience*, *14*(7), 4109–4124. https://doi.org/10.1523/JNEUROSCI.14-07-04109.1994

Chen, G., Lu, H. D., & Roe, A. W. (2008). A Map for Horizontal Disparity in Monkey V2. *Neuron*, *58*(3), 442–450. https://doi.org/10.1016/j.neuron.2008.02.032

Cicmil, N., Cumming, B. G., Parker, A. J., & Krug, K. (2015). Reward modulates the effect of visual cortical microstimulation on perceptual decisions. *ELife*, *4*. https://doi.org/10.7554/eLife.07832

Clery, S., Cumming, B. G., & Nienborg, H. (2017). Decision-Related Activity in Macaque V2 for Fine Disparity Discrimination Is Not Compatible with Optimal Linear Readout. *The Journal of Neuroscience*, *37*(3), 715–725. https://doi.org/10.1523/JNEUROSCI.2445-16.2016

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, *14*(7), 811–819. https://doi.org/10.1038/nn.2842

Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, *12*(12), 1594–1600. https://doi.org/10.1038/nn.2439

Cohen, M. R., & Maunsell, J. H. R. (2011). Using Neuronal Populations to Study the Mechanisms Underlying Spatial and Feature Attention. *Neuron*, *70*(6), 1192–1204. https://doi.org/10.1016/j.neuron.2011.04.029

Cohen, M. R., & Newsome, W. T. (2004). What electrical microstimulation has revealed about the neural basis of cognition. *Current Opinion in Neurobiology*, *14*(2), 169–177. https://doi.org/10.1016/j.conb.2004.03.016

Cohen, M. R., & Newsome, W. T. (2008). Context-Dependent Changes in Functional Circuitry in Visual Area MT. *Neuron*, *60*(1), 162–173. https://doi.org/10.1016/j.neuron.2008.08.007

Coull, J. T., Frith, C. D., Büchel, C., & Nobre, A. C. (2000). Orienting attention in time: Behavioural and neuroanatomical distinction between exogenous and endogenous shifts. *Neuropsychologia*, *38*(6), 808–819. https://doi.org/10.1016/S0028-3932(99)00132-3

Crapse, T. B., & Basso, M. A. (2015a). Insights into decision making using choice probability. *Journal of Neurophysiology*, *114*(6), 3039–3049. https://doi.org/10.1152/jn.00335.2015

Cumming, B. G., & DeAngelis, G. C. (2001). The Physiology of Stereopsis. *Annual Review of Neuroscience*, *24*(1), 203–238. https://doi.org/10.1146/annurev.neuro.24.1.203

Cumming, B. G., & Nienborg, H. (2016). Feedforward and feedback sources of choice probability in neural population responses. *Current Opinion in Neurobiology*, *37*, 126–132. https://doi.org/10.1016/j.conb.2016.01.009

DeAngelis, G. C. (2000). Seeing in three dimensions: The neurophysiology of stereopsis. *Trends in Cognitive Sciences*, *4*(3), 80–90. https://doi.org/10.1016/S1364-6613(99)01443-6

DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature*, *394*(6694), 677–680. https://doi.org/10.1038/29299

DeAngelis, G. C., & Newsome, W. T. (1999a). Organization of Disparity-Selective Neurons in Macaque Area MT. *The Journal of Neuroscience*, *19*(4), 1398–1415. https://doi.org/10.1523/JNEUROSCI.19-04-01398.1999

DeAngelis, G. C., & Newsome, W. T. (1999b). Organization of Disparity-Selective Neurons in Macaque Area MT. *The Journal of Neuroscience*, *19*(4), 1398–1415. https://doi.org/10.1523/JNEUROSCI.19-04-01398.1999

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Desimone, R., & Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *The Journal of Comparative Neurology*, *248*(2), 164–189. https://doi.org/10.1002/cne.902480203

DeYoe, E. A., & Van Essen, D. C. (1985). Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature*, *317*(6032), 58–61. https://doi.org/10.1038/317058a0

Dodd, J. V., Krug, K., Cumming, B. G., & Parker, A. J. (2001). Perceptually Bistable Three-Dimensional Figures Evoke High Choice Probabilities in Cortical Area MT. *The Journal of Neuroscience*, *21*(13), 4809–4821. https://doi.org/10.1523/JNEUROSCI.21-13-04809.2001

Dotson, N. M., Hoffman, S. J., Goodell, B., & Gray, C. M. (2018). Feature-Based Visual Short-Term Memory Is Widely Distributed and Hierarchically Organized. *Neuron*, *99*(1), 215-226.e4. https://doi.org/10.1016/j.neuron.2018.05.026

Eastman, K. M., & Huk, A. C. (2012). PLDAPS: A Hardware Architecture and Software Toolbox for Neurophysiology Requiring Complex Visual Stimuli and Online Behavioral Control. *Frontiers in Neuroinformatics*, *6*. https://doi.org/10.3389/fninf.2012.00001

Federer, F., Ta'afua, S., Merlin, S., Hassanpour, M. S., & Angelucci, A. (2021). Stream-specific feedback inputs to the primate primary visual cortex. *Nature Communications*, *12*(1), 228. https://doi.org/10.1038/s41467-020-20505-5

Felleman, D. J., Burkhalter, A., & Essen, D. C. V. (1997). Cortical connections of areas V3 and VP of macaque monkey extrastriate visual cortex. *Journal of Comparative Neurology*, *379*(1), 21–47. https://doi.org/10.1002/(SICI)1096-9861(19970303)379:1<21::AID-CNE3>3.0.CO;2-K

Fries, P., Womelsdorf, T., Oostenveld, R., & Desimone, R. (2008). The Effects of Visual Stimulation and Selective Visual Attention on Rhythmic Neuronal Synchronization in Macaque Area V4. *Journal of Neuroscience*, *28*(18), 4823–4835. https://doi.org/10.1523/JNEUROSCI.4499-07.2008

Galvan, A., Stauffer, W. R., Acker, L., El-Shamayleh, Y., Inoue, K., Ohayon, S., & Schmid, M. C. (2017). Nonhuman Primate Optogenetics: Recent Advances and

Future Directions. *The Journal of Neuroscience*, *37*(45), 10894–10903. https://doi.org/10.1523/JNEUROSCI.1839-17.2017

Gonzalez, F., & Perez, R. (1998). Neural mechanisms underlying stereoscopic vision. *Progress in Neurobiology*, *55*(3), 191–224. https://doi.org/10.1016/S0301-0082(98)00012-4

Goris, R. L. T., Ziemba, C. M., Stine, G. M., Simoncelli, E. P., & Movshon, J. A. (2017). Dissociation of Choice Formation and Choice-Correlated Activity in Macaque Visual Cortex. *The Journal of Neuroscience*, *37*(20), 5195–5203. https://doi.org/10.1523/JNEUROSCI.3331-16.2017

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* (pp. xi, 455). John Wiley.

Haefner, R. M., Berkes, P., & Fiser, J. (2016). Perceptual Decision-Making as Probabilistic Inference by Neural Sampling. *Neuron*, *90*(3), 649–660. https://doi.org/10.1016/j.neuron.2016.03.020

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

Haimerl, C., Ruff, D. A., Cohen, M. R., Savin, C., & Simoncelli, E. P. (2021). *Targeted comodulation supports flexible and accurate decoding in V1* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.02.23.432351

Haimerl, C., Savin, C., & Simoncelli, E. (2019). Flexible information routing in neural populations through stochastic comodulation. *Advances in Neural Information Processing Systems*, *33*.

Harris, K. D., & Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature*, *503*(7474), 51–58. https://doi.org/10.1038/nature12654

Hastie, T. J., Tibshirani, S., & Friedman, H. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Helmholtz, H. von. (1867). *Handbuch der physiologischen Optik*. Voss.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3), 574–591. https://doi.org/10.1113/jphysiol.1959.sp006308

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Hubel, D. H., Wiesel, T. N., Yeagle, E. M., Lafer-Sousa, R., & Conway, B. R. (2015). Binocular Stereoscopy in Visual Areas V-2, V-3, and V-3A of the Macaque Monkey. *Cerebral Cortex*, *25*(4), 959–971. https://doi.org/10.1093/cercor/bht288

Hubel, D., & Livingstone, M. (1987). Segregation of form, color, and stereopsis in primate area 18. *The Journal of Neuroscience*, *7*(11), 3378–3415. https://doi.org/10.1523/JNEUROSCI.07-11-03378.1987

Huk, A. C., Katz, L. N., & Yates, J. L. (2017). The Role of the Lateral Intraparietal Area in (the Study of) Decision Making. *Annual Review of Neuroscience*, *40*(1), 349–372. https://doi.org/10.1146/annurev-neuro-072116-031508

Kanitscheider, I., Coen-Cagli, R., & Pouget, A. (2015). Origin of information-limiting noise correlations. *Proceedings of the National Academy of Sciences*, *112*(50), E6973–E6982. https://doi.org/10.1073/pnas.1508738112

Kawaguchi, K., Clery, S., Pourriahi, P., Seillier, L., Haefner, R. M., & Nienborg, H. (2018). Differentiating between Models of Perceptual Decision Making Using Pupil Size Inferred Confidence. *The Journal of Neuroscience*, *38*(41), 8874–8888. https://doi.org/10.1523/JNEUROSCI.0735-18.2018

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(ECVP Abstract Supplement), 89.

Kohn, A., & Smith, M. A. (2005). Stimulus Dependence of Neuronal Correlation in Primary Visual Cortex of the Macaque. *Journal of Neuroscience*, *25*(14), 3661–3673. https://doi.org/10.1523/JNEUROSCI.5106-04.2005

Krishna, A., Tanabe, S., & Kohn, A. (2020). Decision Signals in the Local Field Potentials of Early and Mid-Level Macaque Visual Cortex. *Cerebral Cortex*, 1–15. https://doi.org/10.1093/cercor/bhaa218

Krug, K. (2004). A common neuronal code for perceptual processes in visual cortex? Comparing choice and attentional correlates in V5/MT. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1446), 929–941. https://doi.org/10.1098/rstb.2003.1415

Krug, K., Cicmil, N., Parker, A. J., & Cumming, B. G. (2013). A Causal Role for V5/MT Neurons Coding Motion-Disparity Conjunctions in Resolving Perceptual Ambiguity. *Current Biology*, *23*(15), 1454–1459. https://doi.org/10.1016/j.cub.2013.06.023

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434. https://doi.org/10.1364/JOSAA.20.001434

Li, W., Piëch, V., & Gilbert, C. D. (2008). Learning to Link Visual Contours. *Neuron*, *57*(3), 442–451. https://doi.org/10.1016/j.neuron.2007.12.011

Livingstone, M. S., & Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, *4*(1), 309–356. https://doi.org/10.1523/JNEUROSCI.04-01-00309.1984

Lueckmann, J.-M., Macke, J. H., & Nienborg, H. (2018). Can Serial Dependencies in Choices and Neural Activity Explain Choice Probabilities? *The Journal of Neuroscience*, *38*(14), 3495–3506. https://doi.org/10.1523/JNEUROSCI.2225-17.2018

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, *49*(5), 1148–1167. https://doi.org/10.1152/jn.1983.49.5.1148

Maunsell, J., & van Essen, D. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *The Journal of Neuroscience*, *3*(12), 2563–2586. https://doi.org/10.1523/JNEUROSCI.03-12-02563.1983

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of Attention on Orientation-Tuning Functions of Single Neurons in Macaque Cortical Area V4. *The Journal of*

Neuroscience, *19*(1), 431–441. https://doi.org/10.1523/JNEUROSCI.19-01-00431.1999

McAdams, C. J., & Maunsell, J. H. R. (2000). Attention to Both Space and Feature Modulates Neuronal Responses in Macaque Area V4. *Journal of Neurophysiology*, *83*(3), 1751–1755. https://doi.org/10.1152/jn.2000.83.3.1751

McFarland, J. M., Cui, Y., & Butts, D. A. (2013). Inferring Nonlinear Neuronal Computation Based on Physiologically Plausible Inputs. *PLoS Computational Biology*, *9*(7), e1003143. https://doi.org/10.1371/journal.pcbi.1003143

Mirabella, G., Bertini, G., & Chelazzi, L. (2007). Neurons in Area V4 of the Macaque Translate Attended Visual Features into Behaviorally Relevant Categories. *Neuron*, *54*, 303–318. https://doi.org/10.1016/j.neuron.2007.04.007

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2009). Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. *Neuron*, *63*(6), 879–888. https://doi.org/10.1016/j.neuron.2009.09.013

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, *421*(6921), 370–373. https://doi.org/10.1038/nature01341

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784. https://doi.org/10.1126/science.4023713

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., & Pouget, A. (2014). Information-limiting correlations. *Nature Neuroscience*, *17*(10), 1410–1417. https://doi.org/10.1038/nn.3807

Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S., & Churchland, A. K. (2019). Single-trial neural dynamics are dominated by richly varied movements. *Nature Neuroscience*, *22*(10), 1677–1686. https://doi.org/10.1038/s41593-019-0502-4

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, *401*(6754), 695–698. https://doi.org/10.1038/44409

Newsome, W., & Pare, E. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *The Journal of Neuroscience*, *8*(6), 2201–2211. https://doi.org/10.1523/JNEUROSCI.08-06-02201.1988

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989a). Neuronal correlates of a perceptual decision. *Nature*, *341*(6237), 52–54. https://doi.org/10.1038/341052a0

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989b). Neuronal correlates of a perceptual decision. *Nature*, *341*(6237), 52–54. https://doi.org/10.1038/341052a0

Newsome, W. T., Mikami, A., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion. *Journal of Neurophysiology*, *55*(6), 1340–1351. https://doi.org/10.1152/jn.1986.55.6.1340

Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J., & Cohen, M. R. (2018). Learning and attention reveal a general relationship between population activity and behavior. *Science*, *359*(6374), 463–465. https://doi.org/10.1126/science.aao0284

Nienborg, H., & Cumming, B. (2010). Correlations between the activity of sensory neurons and behavior: How much do they tell us about a neuron's causality?

Current Opinion in Neurobiology, *20*(3), 376–381. https://doi.org/10.1016/j.conb.2010.05.002

Nienborg, H., & Cumming, B. G. (2006). Macaque V2 Neurons, But Not V1 Neurons, Show Choice-Related Activity. *Journal of Neuroscience*, *26*(37), 9567–9578. https://doi.org/10.1523/JNEUROSCI.2256-06.2006

Nienborg, H., & Cumming, B. G. (2007). Psychophysically measured task strategy for disparity discrimination is reflected in V2 neurons. *Nature Neuroscience*, *10*(12), 1608–1614. https://doi.org/10.1038/nn1991

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

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

Nienborg, H., & Cumming, B. G. (2014). Decision-Related Activity in Sensory Neurons May Depend on the Columnar Architecture of Cerebral Cortex. *Journal of Neuroscience*, *34*(10), 3579–3585. https://doi.org/10.1523/JNEUROSCI.2340-13.2014

Nienborg, H., R. Cohen, M., & Cumming, B. G. (2012). Decision-Related Activity in Sensory Neurons: Correlations Among Neurons and with Behavior. *Annual Review of Neuroscience*, *35*(1), 463–483. https://doi.org/10.1146/annurev-neuro-062111-150403

Nienborg, H., & Roelfsema, P. R. (2015). Belief states as a framework to explain extra-retinal influences in visual cortex. *Current Opinion in Neurobiology*, *32*, 45–52. https://doi.org/10.1016/j.conb.2014.10.013

Orban, G. A., Saunders, R. C., & Vandenbussche, E. (1995). Lesions of the Superior Temporal Cortical Motion Areas Impair Speed Discrimination in the Macaque Monkey. *European Journal of Neuroscience*, *7*(11), 2261–2276. https://doi.org/10.1111/j.1460-9568.1995.tb00647.x

Parker, A. J., & Newsome, W. T. (1998). SENSE AND THE SINGLE NEURON: Probing the Physiology of Perception. *Annual Review of Neuroscience*, *21*(1), 227–277. https://doi.org/10.1146/annurev.neuro.21.1.227

Pesaran, B., Vinck, M., Einevoll, G. T., Sirota, A., Fries, P., Siegel, M., Truccolo, W., Schroeder, C. E., & Srinivasan, R. (2018). Investigating large-scale brain dynamics using field potential recordings: Analysis and interpretation. *Nature Neuroscience*, *21*(7), 903–919. https://doi.org/10.1038/s41593-018-0171-8

Petro, L. S., Vizioli, L., & Muckli, L. (2014). Contributions of cortical feedback to sensory processing in primary visual cortex. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01223

Pitkow, X., Liu, S., Angelaki, D. E., DeAngelis, G. C., & Pouget, A. (2015). How Can Single Sensory Neurons Predict Behavior? *Neuron*, *87*(2), 411–423. https://doi.org/10.1016/j.neuron.2015.06.033

Poggio, G. F., & Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology*, *40*(6), 1392–1405. https://doi.org/10.1152/jn.1977.40.6.1392

Poggio, G., Gonzalez, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity selectivity. *The Journal of Neuroscience*, *8*(12), 4531–4550. https://doi.org/10.1523/JNEUROSCI.08-12-04531.1988

Ponce, C. R., Lomber, S. G., & Born, R. T. (2008). Integrating motion and depth via parallel pathways. *Nature Neuroscience*, *11*(2), 216–223. https://doi.org/10.1038/nn2039

Prince, S. J. D., Pointon, A. D., Cumming, B. G., & Parker, A. J. (2002). Quantitative Analysis of the Responses of V1 Neurons to Horizontal Disparity in Dynamic Random-Dot Stereograms. *Journal of Neurophysiology*, *87*(1), 191–208. https://doi.org/10.1152/jn.00465.2000

Quinn, K. R., Seillier, L., Butts, D. A., & Nienborg, H. (2021). Decision-related feedback in visual cortex lacks spatial selectivity. *Nature Communications*, *12*(1), 4473. https://doi.org/10.1038/s41467-021-24629-0

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Reynolds, J. H., & Chelazzi, L. (2004). ATTENTIONAL MODULATION OF VISUAL PROCESSING. *Annual Review of Neuroscience*, *27*(1), 611–647. https://doi.org/10.1146/annurev.neuro.26.041002.131039

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

Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research*, *179*(1), 3–20. https://doi.org/10.1016/0006-8993(79)90485-2

Roe, A. W., Parker, A. J., Born, R. T., & DeAngelis, G. C. (2007). Disparity Channels in Early Vision. *Journal of Neuroscience*, *27*(44), 11820–11831. https://doi.org/10.1523/JNEUROSCI.4164-07.2007

Roelfsema, P. R., & de Lange, F. P. (2016). Early Visual Cortex as a Multiscale Cognitive Blackboard. *Annual Review of Vision Science*, *2*(1), 131–151. https://doi.org/10.1146/annurev-vision-111815-114443

Roelfsema, P. R., van Ooyen, A., & Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends in Cognitive Sciences*, *14*(2), 64–71. https://doi.org/10.1016/j.tics.2009.11.005

Rorie, A. E., Gao, J., McClelland, J. L., & Newsome, W. T. (2010). Integration of Sensory and Reward Information during Perceptual Decision-Making in Lateral Intraparietal Cortex (LIP) of the Macaque Monkey. *PLoS ONE*, *5*(2), e9308. https://doi.org/10.1371/journal.pone.0009308

Ruff, D. A., Ni, A. M., & Cohen, M. R. (2018). Cognition as a Window into Neuronal Population Space. *Annual Review of Neuroscience*, *41*, 77–97.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience*, *5*(7), 631–632. https://doi.org/10.1038/nn876

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

Salzman, C., Murasugi, C., Britten, K., & Newsome, W. (1992). Microstimulation in visual area MT: Effects on direction discrimination performance. *The Journal of Neuroscience*, *12*(6), 2331–2355. https://doi.org/10.1523/JNEUROSCI.12-06-02331.1992

Seillier, L., Lorenz, C., Kawaguchi, K., Ott, T., Nieder, A., Pourriahi, P., & Nienborg, H. (2017). Serotonin Decreases the Gain of Visual Responses in Awake Macaque V1. *The Journal of Neuroscience*, *37*(47), 11390–11405. https://doi.org/10.1523/JNEUROSCI.1339-17.2017

Shadlen, M., Britten, K., Newsome, W., & Movshon, J. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *The Journal of Neuroscience*, *16*(4), 1486–1510. https://doi.org/10.1523/JNEUROSCI.16-04-01486.1996

Shiozaki, H. M., Tanabe, S., Doi, T., & Fujita, I. (2012). Neural Activity in Cortical Area V4 Underlies Fine Disparity Discrimination. *Journal of Neuroscience*, *32*(11), 3830–3841. https://doi.org/10.1523/JNEUROSCI.5083-11.2012

Shipp, S., & Zeki, S. (1985). Segregation of pathways from area V2 to areas V4 and V5 of macaque monkey visual cortex. *Nature*, *315*, 322–324. https://doi.org/10.1038/315322a0

Siegel, M., Buschman, T. J., & Miller, E. K. (2015). Cortical information flow during flexible sensorimotor decisions. *Science*, *348*(6241), 1352–1355. https://doi.org/10.1126/science.aab0551

Spyropoulos, G., Bosman, C. A., & Fries, P. (2018). A theta rhythm in macaque visual cortex and its attentional modulation. *Proceedings of the National Academy of Sciences*, *115*(24), E5614–E5623. https://doi.org/10.1073/pnas.1719433115

Steinmetz, N. A., Koch, C., Harris, K. D., & Carandini, M. (2018). Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. *Current Opinion in Neurobiology*, *50*, 92–100. https://doi.org/10.1016/j.conb.2018.01.009

Steinmetz, N. A., Zatka-Haas, P., Carandini, M., & Harris, K. D. (2019). Distributed coding of choice, action and engagement across the mouse brain. *Nature*, *576*(7786), 266–273. https://doi.org/10.1038/s41586-019-1787-x

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

Talluri, B. C., Kang, I., Lazere, A., Quinn, K. R., Kaliss, N., Yates, J. L., Butts, D. A., & Nienborg, H. (2022). *Activity in primate visual cortex is minimally driven by spontaneous movements* [Preprint]. Neuroscience. https://doi.org/10.1101/2022.09.08.507006

Thomas, O. M., Cumming, B. G., & Parker, A. J. (2002). A specialization for relative disparity in V2. *Nature Neuroscience*, *5*(5), 472–478. https://doi.org/10.1038/nn837

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, *23*(8), 775–785. https://doi.org/10.1016/0042-6989(83)90200-6

Treue, S., & Maunsell, J. H. R. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, *382*(6591), 539–541. https://doi.org/10.1038/382539a0

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Treutwein, B., & Strasburger, H. (1999). Fitting the psychometric function. *Perception & Psychophysics*, *61*(1), 87–106. https://doi.org/10.3758/BF03211951

Ts'o, D. Y., Roe, A. W., & Gilbert, C. D. (2001). A hierarchy of the functional organization for color, form and disparity in primate visual area V2. *Vision Research*, *41*(10–11), 1333–1349. https://doi.org/10.1016/S0042-6989(01)00076-1

Uka, T. (2005). Neural Correlates of Fine Depth Discrimination in Monkey Inferior Temporal Cortex. *Journal of Neuroscience*, *25*(46), 10796–10802. https://doi.org/10.1523/JNEUROSCI.1637-05.2005

Uka, T., & DeAngelis, G. C. (2003). Task-specific contribution of area MT to stereoscopic depth discrimination. *Journal of Vision*, *3*(9), 96–96. https://doi.org/10.1167/3.9.96

Uka, T., & DeAngelis, G. C. (2004). Contribution of Area MT to Stereoscopic Depth Perception: Choice-Related Response Modulations Reflect Task Strategy. *Neuron*, *42*(2), 297–310.

Uka, T., & DeAngelis, G. C. (2006). Linking Neural Representation to Function in Stereoscopic Depth Perception: Roles of the Middle Temporal Area in Coarse versus Fine Disparity Discrimination. *Journal of Neuroscience*, *26*(25), 6791–6802. https://doi.org/10.1523/JNEUROSCI.5435-05.2006

Umeda, K., Tanabe, S., & Fujita, I. (2007). Representation of Stereoscopic Depth Based on Relative Disparity in Macaque Area V4. *Journal of Neurophysiology*, *98*(1), 241–252. https://doi.org/10.1152/jn.01336.2006

van Bergen, R. S., & Kriegeskorte, N. (2020). Going in circles is the way forward: The role of recurrence in visual inference. *ArXiv:2003.12128 [Cs, q-Bio]*. http://arxiv.org/abs/2003.12128

Wasmuht, D. F., Parker, A. J., & Krug, K. (2019). Interneuronal correlations at longer time scales predict decision signals for bistable structure-from-motion perception. *Scientific Reports*, *9*(1), 11449. https://doi.org/10.1038/s41598-019-47786-1

Watanabe, M., Tanaka, H., Uka, T., & Fujita, I. (2002). Disparity-Selective Neurons in Area V4 of Macaque Monkeys. *Journal of Neurophysiology*, *87*(4), 1960–1973. https://doi.org/10.1152/jn.00780.2000

Whiteway, M. R., & Butts, D. A. (2019). The quest for interpretable models of neural population activity. *Current Opinion in Neurobiology*, *58*, 86–93. https://doi.org/10.1016/j.conb.2019.07.004

Wichmann, F. A., & Jäkel, F. (2018). Methods in Psychophysics. In J. T. Wixted (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–42). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119170174.epcn507

Wilming, N., Murphy, P. R., Meyniel, F., & Donner, T. H. (2020). Large-scale dynamics of perceptual decision information across human cortex. *Nature Communications*, *11*(1), 5109. https://doi.org/10.1038/s41467-020-18826-6

Wimmer, K., Compte, A., Roxin, A., Peixoto, D., Renart, A., & de la Rocha, J. (2015). Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area MT. *Nature Communications*, *6*(1), 6177. https://doi.org/10.1038/ncomms7177

Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert Visual Attention Modulates Face-Specific Activity in the Human Fusiform Gyrus: FMRI Study. *Journal of Neurophysiology*, *79*(3), 1574–1578. https://doi.org/10.1152/jn.1998.79.3.1574

Wurtz, R. H. (2009). Recounting the impact of Hubel and Wiesel: Recounting the impact of Hubel and Wiesel. *The Journal of Physiology*, *587*(12), 2817–2823. https://doi.org/10.1113/jphysiol.2009.170209

Wurtz, R. H. (2015). Using perturbations to identify the brain circuits underlying active vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1677), 20140205. https://doi.org/10.1098/rstb.2014.0205

Yu, X., & Gu, Y. (2018). Probing Sensory Readout via Combined Choice-Correlation Measures and Microstimulation Perturbation. *Neuron*. https://doi.org/10.1016/j.neuron.2018.08.034

Zeki, S. M. (1974). Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *The Journal of Physiology*, *236*(3), 549–573. https://doi.org/10.1113/jphysiol.1974.sp010452

Zeki, S. M. (1978). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *The Journal of Physiology*, *277*(1), 273–290. https://doi.org/10.1113/jphysiol.1978.sp012272

Zhao, Y., Yates, J. L., Levi, A. J., Huk, A. C., & Park, I. M. (2020). Stimulus-choice (mis)alignment in primate area MT. *PLOS Computational Biology*, *16*(5), e1007614. https://doi.org/10.1371/journal.pcbi.1007614

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, *370*(6485), 140–143. https://doi.org/10.1038/370140a0

9. Appendix

9.1. Table of Figures