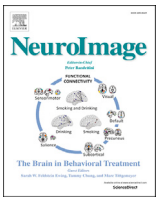
NeuroImage 214 (2020) 116701

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage



Unreliable feedback deteriorates information processing in primary visual cortex

Rekha S. Varrier [a,b], Marcus Rothkirch [a], Heiner Stuke [a], Matthias Guggenmos [a,1,*], Philipp Sterzer [a,b,1]

[a] Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and the Berlin Institute of Health, Chariteplatz 1, 10117, Berlin, Germany

[b] Bernstein Center for Computational Neuroscience, Berlin, Humboldt University Berlin, 10115, Berlin, Germany

ARTICLE INFO

Keywords:
Unreliable feedback
Perceptual inference
fMRI
Multivariate representation
Visual cortex

It is well-established that increased sensory uncertainty impairs perceptual decision-making and leads to degraded neural stimulus representations. Recently, we also showed that providing unreliable feedback to choices leads to changes in perceptual decision-making similar to those of increased stimulus noise: A deterioration in objective task performance, a decrease in subjective confidence and a lower reliance on sensory information for perceptual inference. To investigate the neural basis of such feedback-based changes in perceptual decision-making, in the present study, two groups of healthy human participants (n = 15 each) performed a challenging visual orientation discrimination task while undergoing functional magnetic resonance imaging (fMRI). Critically, one group received reliable feedback regarding their task performance in an intervention phase, whereas the other group correspondingly received unreliable feedback – thereby keeping stimulus information constant. The effects of feedback reliability on performance and stimulus representation in the primary visual cortex (V1) were studied by comparing the pre- and post-intervention test phases between the groups. Compared to participants who received reliable feedback, those receiving unreliable feedback showed a decline in task performance that was paralleled by reduced distinctness of fMRI response patterns in V1. These results show that environmental uncertainty can affect perceptual inference at the earliest cortical processing stages.

1. Introduction

Bayesian models of brain function frame perception as an inferential process, whereby an internal model of the world is used to infer the most probable causes of the sensory data (Friston, 2005; O'Reilly et al., 2012). For such perceptual inference to be adaptive, the uncertainty of sensory data must be taken into account, whereby uncertain sensory information should be given less weight in perceptual inference (Knill and Pouget, 2004; Adams et al., 2013). In experimental studies, uncertainty of the sensory information is typically manipulated by adding varying levels of noise, which has been shown to result in less informative neural stimulus representations in sensory cortical areas (Hebart et al., 2012; Ludwig et al., 2016; Darcy et al., 2019). In addition to such stimulus-based sensory uncertainty, the weighting of sensory information in perceptual inference also depends on stimulus-independent factors. For instance, it is well-established that expectations about upcoming stimuli can reduce the relative weight given to the sensory input (Sterzer et al., 2008; Summerfield et al., 2008; Series and Seitz, 2013). Recently, we showed that providing unreliable feedback to perceptual decisions likewise leads to a down-weighting of sensory information, paralleled by an up-weighting of prior beliefs (Varrier et al., 2019). This distortion of perceptual inference was also reflected in a deterioration of objective task performance similar to previous work (Herzog and Fahle, 1997, 1999) and a decrease in the subjective confidence about perceptual decisions, just as observed with increases of stimulus-based sensory uncertainty (Mamassian, 2016). Computational modelling further showed that these behavioural effects were well-explained by a model in which unreliable feedback altered the likelihood distributions of the observer, in turn leading to a down-weighting of sensory information (Varrier et al., 2019).

Abbreviations: CW/ CCW, clockwise/counter-clockwise; CV-MANOVA, cross-validated multivariate analysis of variance.

* Corresponding author. Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and the Berlin Institute of Health, Chariteplatz 1, 10117, Berlin, Germany.

E-mail address: matthias.guggenmos@charite.de (M. Guggenmos).

[1] Equal contribution.

https://doi.org/10.1016/j.neuroimage.2020.116701
Received 13 November 2019; Received in revised form 1 February 2020; Accepted 29 February 2020
Available online 3 March 2020

The present study aimed at identifying the neural basis of such feedback-induced changes in perceptual inference. We used functional magnetic resonance imaging (fMRI) to investigate whether the previously reported deterioration of perceptual task performance under unreliable feedback is related to altered neural stimulus representations in sensory cortical areas, similar to those observed in relation to stimulus-based sensory uncertainty. Based on previous fMRI studies showing successful decoding of visual grating orientations from activation patterns in primary visual cortex (V1) (Kamitani and Tong, 2005; Haynes and Rees, 2005; Kok et al., 2012), we employed an orientation discrimination task and examined the distinctness of V1 activation patterns (Allefeld and Haynes, 2014) in relation to different stimulus orientations (Fig. 1). The primary visual cortex or V1 is most sensitive to the low-level properties such as contrasts and spatial frequency (Boynton et al., 1999; Avidan et al., 2002; Tong et al., 2012) that are critical to the orientated gratings used in this experiment, and hence our hypothesis was tested in V1, where the effects on sensory representations were expected to be maximal. The participants' task was to discriminate between clockwise (CW) and counter-clockwise (CCW) deviations of each presented grating from an implicit diagonal. Task difficulty was individually adjusted through a staircase procedure for each participant, and the stimulus-response mapping was randomised using response cues. The main experiment comprised a pre-intervention test phase, a feedback intervention phase and a post-intervention test phase. Critically, the same orientations were presented in all three phases of the experiment, so that any observed changes in behavioural performance or neural pattern representations were independent of physical stimulus properties. Thirty participants were randomly assigned to one of two experimental groups that differed only with respect to the intervention phase: In one group, trial-wise feedback on task performance in this phase was randomised and therefore unreliable, while in the other group feedback was always correct and therefore reliable. The effects of unreliable vs. reliable feedback on task performance and V1 pattern distinctness were assessed by comparing changes from pre-to post-intervention test runs between the two groups.

2. Materials and Methods

2.1. Participants

The study was approved by the ethics committee at Charité - Universitätsmedizin Berlin, and informed consents were collected in writing from all participants. Participants were students from Humboldt University and Charité – Universitätsmedizin Berlin. Thirty-two healthy participants took part in the experiment (18–35 years, mean age = 24.9, 13 female). Of these, two participants were excluded from the experiment before the intervention runs due to chance-level performance in multiple runs in the pre-intervention phase. FMRI data from an additional participant in the reliable feedback group were discarded due to excessive head motion during the post-intervention runs. Further, fMRI data of two of the eight runs in the post-intervention phase of eight participants (four each from the reliable and unreliable feedback groups) were lost due to an error in the scanner sequence. In these participants, the corresponding two runs from the pre-intervention runs were likewise excluded for fMRI data analysis so as to make the pre- and post-intervention data comparable in terms of statistical power for the multivariate analyses.

2.2. General design

To study effects of feedback, participants were divided into two groups which received either reliable (n = 15) or unreliable (n = 15) feedback during a challenging orientation discrimination task. To measure neural responses during stimulus presentation, participants underwent fMRI scanning. Each session consisted of task training outside the scanner, threshold estimation, the main experiment and the functional localiser. The main experiment consisted of three phases: (1) a pre-intervention test phase without feedback, (2) an intervention phase with either reliable or unreliable feedback and (3) a post-intervention test phase without feedback (Fig. 1a). For the three phases of the main experiment and the functional localiser which followed it, stimuli were

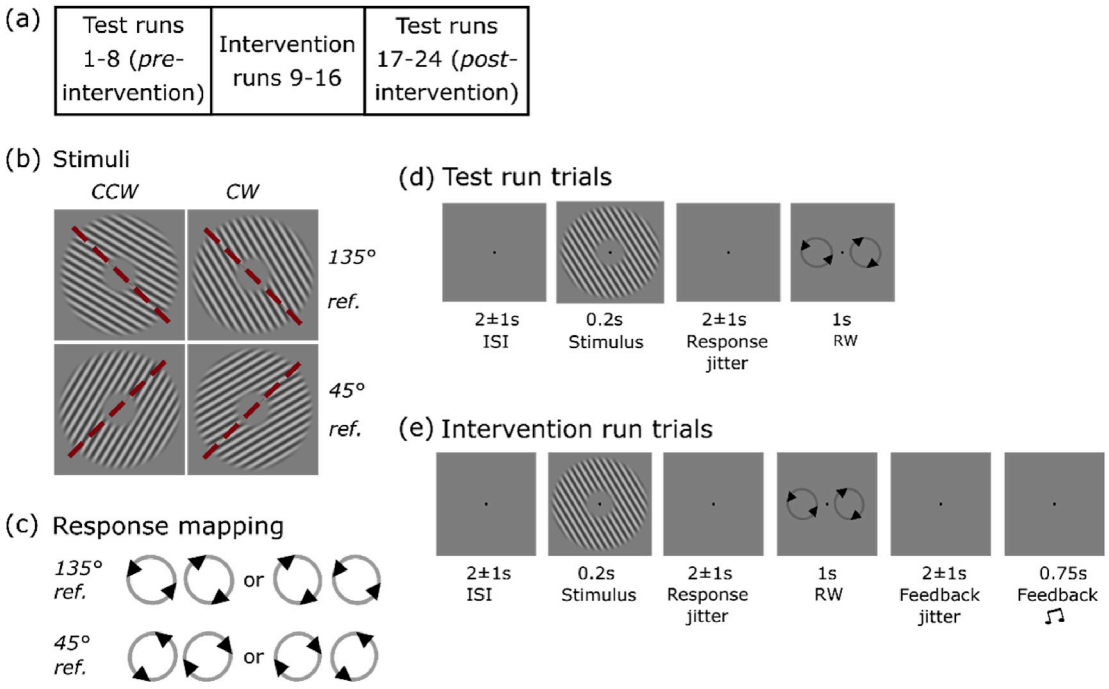


Fig. 1. (a) Design of the main experiment. (b) Sinusoidal grating stimuli of the orientation discrimination task. There were two pairs of stimuli corresponding to implicit diagonals at 135° and 45°. Reference lines are shown in red for illustrative purposes. The depicted orientation deviations from diagonals are exaggerated for illustration. (c) Two types of response mappings assigned CCW (CW) responses either to the left (right) button or the right (left) button of a response box. Time courses of trials in (d) test runs and (e) intervention runs. The response cues are enlarged in (d) and (e) for better visibility. ISI = inter-stimulus interval, RW = response window.

presented at the threshold estimated in the previous step. Feedback-induced changes in behavioural and neural responses were studied by comparing pre- and post-intervention phases in which the visual stimulation remained the same.

2.3. Stimuli

Stimulus presentation was implemented using PsychToolbox 3.0.11 (psychtoolbox.org) for Matlab (MathWorks Inc.). Visual stimuli were presented on a monitor (resolution: 1024 × 768 pixels) and projected using an oblique mirror into the eyes of participants lying in a supine position (total distance 154 cm). The stimuli were high-contrast annular sinusoidal gratings (inner radius = 1.32°, outer radius = 6.69°, spatial frequency = 1.29cpd) with luminance ranging from 25% to 75% of the maximum luminance of the screen. To reduce visual responses to the inner and outer edges of the annuli, these edges were blurred using circular Gaussian functions centred at the inner (1.38°) and outer (6.64°) edges such that the contrast gradually faded until it matched the gray-scale background (standard deviation $\sigma = 0.44°$). This procedure rendered high contrast sinusoidal gratings with soft edges (see Fig. 1b). Image properties other than orientation - such as contrast, spatial frequency and size were identical for all participants and across the whole experiment. To reduce neural adaptation and to avoid "point-of-reference" strategies by the participants, Gabor patches were presented with variable phase shifts in different trials, randomly drawn from 16 equally spaced shifts between 0 and 2π. Grating orientations deviated CW or CCW from the diagonal references (45° or 135°) by an angle defined individually for each participant in the threshold estimation step *prior to* the main experiment (Fig. 1b).

2.4. Responses

In order to orthogonalise stimuli and responses, so as to prevent motor planning during stimulus presentation, participants were informed of the stimulus-response mapping (i.e., which button to press for perceived CCW/CW orientations) only after the stimulus disappeared (Kahnt et al., 2011; Hebart et al., 2012). Response cues were presented for a time window of 1s, during which participants were asked to respond (Fig. 1d–e). The cues were small circles with arrows indicating the response mapping (CCW/CW) and were presented to the left and right of the fixation dot with the inner arcs at 0.48° (visual angle). There were two pairs of response cues corresponding to stimuli with 45° and 135° references, and these could be presented in two sequences (left-right or right-left). All possible presentation sequences of the response cues are illustrated in Fig. 1c.

2.5. Feedback

Trial-by-trial feedback was delivered by means of auditory tones after each response using speakers (Fig. 1e). Before the main experiment, participants familiarised themselves with the tones and their associated meaning during training runs outside the scanner. Feedback tones were audible above the scanner noise, confirmed by asking participants during the scans and in the post-experimental debriefing. The tones were positive, negative or neutral, depending on whether the response was correct, incorrect or missed, respectively. Participants in the unreliable feedback group received pseudo-randomised feedback in the intervention runs, such that in half of the trials, the feedback delivered was faulty (i.e., positive tones after incorrect button presses and negative tones after correct button presses). Participants in the reliable feedback group only received valid feedback.

2.6. Task

On each trial, one of the four stimuli – gratings rotated CW or CCW from an implicit diagonal (45° or 135°, see Fig. 1b) – could be presented, and the participants' task was to report their perceived orientation of the grating as either CCW- or CW-rotated from an implicit diagonal reference. Participants were trained to identify the relevant diagonal for each type of stimulus, which was in the same quadrant as the presented stimulus (for example, if the presented stimulus had an orientation of 60°, the correct percept would be that it was counter-clockwise, since the implicit diagonal in *that* quadrant, i.e., the 45° diagonal, was to be used as the reference).

2.7. Trials

The time course of test and intervention trials are shown in Fig. 1d–e. A trial started with the presentation of the fixation dot (radius 0.1° visual angle) for 2±1s, followed by the presentation of the visual stimulus for 0.2s, which was followed by fixation for 2±1s. Next, the response window was presented for 1s, during which the response mapping was indicated using the response cues, and responses were indicated by pressing one of two buttons of an fMRI-compatible button box. In the intervention runs alone, following the response, there was another fixation window (2±1s), following which the auditory feedback was delivered (0.75s). On each trial, the stimulus orientation was one of four types (CCW/CW tilted with respect to the 45° or the 135° diagonal, see Fig. 1b). Stimuli were presented in a pseudo-random fashion across trials as determined prior to each test or intervention phase such that all stimuli were presented an equal number of times within each run.

2.8. Experimental schedule

At the beginning of the experiment and outside of the scanner, participants were trained in the orientation discrimination task using supra-threshold versions of the gratings (20 min). Once inside the scanner, individual orientation discrimination thresholds for the main experiment were determined using a staircase procedure (10 min). 100% valid feedback was delivered both in the training and the threshold estimation steps to facilitate learning of the task and the response mapping. The main experiment consisted of 24 runs (overall 85 min) and was followed by a short functional localiser task (6 min). At the end of the experiment, participants were debriefed. The three parts are explained in more detail below.

2.8.1. Training

Participants performed training runs in a testing room outside the scanner. The first run consisted of supra-threshold stimuli and the participants manually navigated through each stimulus and response screen at their own pace. The second run consisted likewise of supra-threshold stimuli, but trial timings corresponded to those of the main experiment (time course shown in Fig. 1e). If necessary, the second run was repeated until the participant could make responses in the given time and got at least 80% correct responses.

2.8.2. Staircase procedure

Inside the scanner, participants performed a staircase task to set the deviation of stimuli from the diagonal references at which they could discriminate between orientations with moderate difficulty. To this end, we used two two-down, one-up staircases with equal step-sizes up and down which arrived at an 80% performance threshold. The first phase of the staircase procedure determined the approximate signal threshold and had larger step-sizes (angular deviations from the two diagonals were multiplied by the factors $10^{-0.03}$ and $10^{0.03}$ to decrease or increase thresholds, respectively). The second phase started at the threshold estimated by the first staircase and used a fixed step size of 0.3°, both for an increase and for a decrease in deviations from the diagonal. Each staircase stopped when a certain number of reversals (six reversals for phase one, ten reversals for phase two) or 80 trials were reached. Thresholds were estimated by averaging the last four and six reversal points in the staircase phases one and two, respectively.

2.8.3. Main experiment

The main experiment consisted of 24 runs, split into three parts (Fig. 1a): Runs 1–8 were pre-intervention test runs (without feedback), runs 9–16 were intervention runs (with feedback) and runs 17–24 were post-intervention test runs (without feedback). Each run consisted of 32 trials, in which each of the four stimulus conditions (CCW/CW deviations from the 45° and 135° diagonal references) was shown eight times. In the intervention runs, half of the participants received reliable and the other half received unreliable feedback. The pre- and post-intervention test runs were identical in structure, and their purpose was to measure changes induced during the feedback intervention.

2.8.4. Functional localiser

Lastly, a functional localiser was run in which the four stimulus conditions of the main experiment were presented along with a fixation-only baseline condition. The five conditions were shown in blocks of 12-s duration and were repeated six times in a pseudo-random order. During the 12-s presentations of the four stimuli, all 16 phase shifts were shown in a random order at a rate of 3.33Hz. To ensure that participants fixated during the functional localiser run, a central fixation dot was present in all conditions and changed its colour to red briefly at random (0.3s), and participants had to press the left response button to indicate whenever this event occurred.

2.8.5. Debriefing

At the end of the experiment, all participants were given questionnaires to probe their awareness of having received unreliable feedback and motivation to do the task. To understand participants' awareness of feedback manipulation, they rated their percentage reliability on feedback, the degree to which they suspected feedback manipulation and the intervention run number at which their trust in feedback changed. Participants also rated their motivation to do the task in each phase of the experiment as a percentage value. These questions are quoted verbatim in the Supplementary Methods.

2.9. Eye-tracking

To ensure fixation, an MRI-compatible video-based eye-tracker (iView XTM MRI 50Hz, SensoMotoric Instruments, Teltow, Germany) was used to monitor participants' gaze position throughout the experiment. Eye-tracking data could not be collected from two participants due to difficulties in calibration or in the detection of pupil and corneal reflex by the camera. In all other participants, partial or full data were collected and pre-processed in the following steps: (1) removal of invalid data points, (2) cubic-spline interpolation of missing data points when there was fewer than eight missing data points (160 ms), (3) removal of linear trends and (4) high-pass filtering of data by computing running averages across five consecutive points (100 ms). After pre-processing, data corresponding to the stimulus presentation windows (200 ms) were extracted. Next, participants with a high proportion of missing data (defined as (1) more than 70% invalid data points overall or within the stimulus presentation windows of the pre- or the post-intervention phases, or (2) no data at all from more than six runs within either the pre- or post-intervention phases) were excluded.

2.10. FMRI data acquisition and processing

2.10.1. Data acquisition

Functional brain images were acquired at a 3T Siemens Trio (Erlangen, Germany) scanner using a gradient echo-planar imaging sequence and a 12-channel head coil. Each run in the pre- and post-intervention phases consisted of 90 T2*-weighted whole-brain volumes each, and the functional localiser run consisted of 180 whole-brain volumes. Other parameters remained the same during the main experiment and the localiser (TR = 2s, TE = 30 ms, flip angle = 78°, 33 slices, descending acquisition, 3 mm isotropic resolution, 0.7 mm gap between slices). In addition, high-resolution structural T1-weighted images were acquired using the MPRAGE sequence (TR = 1.9s, TE = 2.52 ms, flip angle = 9°, 192 slices, 1 mm isotropic resolution).

2.10.2. FMRI data processing

The functional images were corrected for slice acquisition delays and translational/rotational motion using the MATLAB-based Statistical Parametric Mapping Toolbox (SPM12, www.fil.ion.ucl.ac.uk/spm). Next, the functional images and the probabilistic V1 mask from the SPM-based Anatomy Toolbox (Eickhoff et al., 2005) were aligned in the following steps: (1) the V1 mask and the Colin27 single-subject brain template (which was in alignment with the anatomical V1 mask) were mapped to subject space, and (2) functional images in their native space were co-registered and resliced to match Colin27 (and consequently the V1 mask). Following co-registration, the functional images were smoothed with a 3 mm (FWHM) Gaussian kernel (Op de Beeck, 2010; Misaki et al., 2013; Gardumi et al., 2016; Hendriks et al., 2017). Next, three separate general linear models (GLMs) were defined for each participant: One GLM for the pre-intervention runs, one for the post-intervention runs, and one for the functional localiser runs. In each GLM, the four stimuli were included as separate regressors, and a fifth regressor encoded the response windows with button presses. These regressors were then convolved with the canonical haemodynamic response function as implemented in SPM12. In addition, the six translation and rotation parameters obtained from the motion correction step were included in each model as regressors of no interest. The GLMs from the pre- and post-intervention runs were used to estimate pattern distinctness, and the GLM from the functional localiser was used to create individual participant-specific regions of interest (ROI) based on t-contrast maps of voxels that responded to the visual stimuli irrespective of orientation and the probabilistic V1 mask. Voxels within brain area V1 were selected for the ROI if they (1) had a probability of greater than 50% of belonging to V1 (referred to as area hOc1 in the SPM-based Anatomy Toolbox (Eickhoff et al., 2005)) and (2) were significant at an uncorrected t-contrast threshold of 0.05. This voxel selection process yielded ROIs with comparable numbers of V1 voxels in both groups of participants (unreliable feedback: 225.4 ± 13.51, reliable feedback: 219.25 ± 17.35 voxels; two-tailed, two-sample *t*-test: $t(27) = 0.28$, $p = .78$). All analyses of pattern distinctness and average responses (see below) were based on these ROIs.

2.10.3. Estimation of pattern distinctness

To estimate the distinctness of stimulus-evoked activation patterns in V1, we used the cross-validated (CV)-MANOVA algorithm (Allefeld and Haynes, 2014). CV-MANOVA performs a leave-one-run-out cross-validation to compute an unbiased estimate for the distinctness of activation patterns. When used to compare two multivariate patterns like in our study, this measure of pattern distinctness is analogous to the Mahalanobis distance. Here the pattern distinctness thus corresponded to the cross-validated Mahalanobis distance between Stimuli with the same diagonal reference. By averaging across the distinctness estimates corresponding to the two stimulus pairs, we obtained a single estimate of pattern distinctness for each participant and test phase (pre-/post-intervention).

2.11. Statistical analyses

The key dependent variables to test our hypotheses about the effects of unreliable feedback were (1) the orientation discrimination performance within each test phase (pre-/post-intervention), quantified as the mean percentage of correct responses averaged across all the test runs within a phase, and (2) the distinctness of patterns in V1 computed for each test phase as described in the previous section.

2 × 2 mixed-design ANOVAs were performed separately for task performance and pattern distinctness. The analyses are called "mixed-design", because there were two factors in the analysis, one of which

was between-subject (feedback type: unreliable or reliable) and the other within-subject (test phase: pre-/post-intervention) . The orientation discrimination threshold was used as a covariate of no interest. Our critical prediction was a significant interaction between feedback type and test phase. In case of significant interactions, post-hoc one-sample *t*-tests (two-tailed) of changes in performance and pattern distinctness were performed to further understand the nature of this interaction. To determine the effect sizes of the behavioural and neural changes, Cohen's d was estimated separately for changes in performance and pattern distinctness between the two groups (unreliable/reliable feedback). These estimates were corrected for the small sample size (Durlak, 2009).

Since we hypothesised that unreliable feedback impairs both task performance and stimulus representations, we tested whether the changes in performance ($\Delta_{\text{Percent correct}}$) and pattern distinctness ($\Delta_{\text{Pattern distinctness}}$) correlated with each other, using the Robust Correlation Toolbox (Pernet et al., 2013). Please note that an analogous analysis was not performed for the reliable feedback group, because the aim of the correlation analysis was to specifically test for a systematic change in the variances induced by unreliable feedback and not the neural correlates of perceptual learning (which we could in principle have probed by correlating behaviour with neural signals in the reliable feedback group). In addition to our hypothesis-driven ROI analysis of V1, an exploratory whole-brain searchlight was also performed post-hoc to study the changes in information representation in other brain areas. The details of this analysis are given in the Supplementary Information.

To study changes in overall arousal and attention as a result of unreliable feedback, the changes in overall neural activity in the selected voxels in V1 was analysed post-hoc using a 2 × 2 mixed-design ANOVA with the same factors and covariate as used to study the changes in pattern distinctness. The mean beta estimates for responses to all stimuli (irrespective of orientation) were computed independently for both test phases (pre/post-intervention) by averaging across all voxels within the individually defined V1 ROIs (see above), and this was used as the dependent variable.

Lastly, we performed post-hoc control analyses to understand if non-perceptual mechanisms could have influenced the changes in performance or pattern distinctness. First of all, the eye-tracking data were analysed to test whether fixation changed as a result of unreliable feedback and whether stimulus orientations (CW vs. CCW) could be decoded from the eye-tracking data (Thielen et al., 2019). Next, the subjective responses to the debriefing questionnaire were analysed to (1) probe the degree of awareness of feedback manipulation and its influence on the observed effects in the unreliable feedback group and (2) investigate if changes in motivation ratings paralleled the main results on the behavioural/neuroimaging data. The performance in the colour-change detection task performed during the functional localiser run was also compared between the two groups. The details of these analyses are described in the Supplementary Information section.

Finally, we performed two additional analyses to further examine the role of potential confounds: (1) To examine the strength of the *absent* interaction effects and correlations, we performed Bayesian analyses using the MATLAB package bayesFactor (Krekelberg, 2019); and (2) to understand their influence on the main behavioural and neural results, we repeated the corresponding mixed-design ANOVAs with each of the potential confounds (e.g. awareness of feedback manipulation, change in fixation etc.) as an additional covariate.

2.12. Data availability

Whole-brain pattern univariate and multivariate contrast maps (used for the searchlight and mean activity analyses) have been uploaded to NeuroVault (https://identifiers.org/neurovault.collection:6040). Since CV-MANOVA gives out only single-value estimates for the ROI analysis (V1) of pattern distinctness, we have provided the pattern distinctness estimates for each test phase. These, together with the behavioural data and the codes used for data analysis have been uploaded to GitHub (https://github.com/rvarrier/fmri_unreliablefb).

3. Results

In this study, we investigated the effects of unreliable feedback by delivering it in a dedicated *intervention* phaseand subsequently measuring its effects in an ensuing *test* phase in which no feedback was delivered. We hypothesised that unreliable feedback would lead to decreases in task performance and the precision of stimulus representation in the visual cortex. To test this, the percentage of correct responses and neural pattern distinctness in V1 were compared between the pre- and post-intervention test phases, and between participant groups receiving unreliable and reliable feedback. Participants performed an orientation discrimination task where stimuli deviated from diagonal references at individually determined thresholds. On average, these thresholds were comparable between the two groups (unreliable: $M = 7.85°$, $SE = 1.53°$, reliable: $M = 8.08°$, $SE = 0.89°$; two-tailed, two-sample *t*-test: $t(28) = 0.13$, $p = .9$).

3.1. Unreliable feedback impairs task performance

Behaviourally, unreliable feedback during the intervention phase led to a significant decline in discrimination performance, as shown in a two-way mixed ANOVA, where the between-subject factor feedback type (unreliable vs. reliable) and the within-subject factor test phase (pre-vs. post-intervention) showed a significant interaction effect ($F(1,27) = 7.26$, $p = .01$; $\eta_p^2 = 0.21$, Fig. 2). Post-hoc two-tailed one-sample *t*-tests showed a significant decrease in performance after unreliable feedback ($M = -6.88$, $SE = 2.60$, $t(14) = -2.64$, $p = .02$) but not reliable feedback ($M = 3.49$, $SE = 2.73$, $t(14) = 1.28$, $p = .22$). The effect size (Cohen's d) of the change in performance between the two groups was 0.94.

3.2. Unreliable feedback deteriorates neural stimulus representations in V1

Neural effects of unreliable feedback were assessed by estimating the distinctness of activation patterns in stimulus-responsive parts of V1 evoked by CW- vs. CCW-rotated gratings based on fMRI data from 29 participants (unreliable feedback: n = 15, reliable feedback n = 14; see Materials and Methods). In line with our hypothesis, and in striking analogy to the behavioural results, we found a significant interaction of feedback type and test phase ($F(1,26) = 5.98$, $p = .02$, $\eta_p^2 = 0.19$; Fig. 3). Again, post-hoc one-sample *t*-tests (two-tailed) showed that there was a

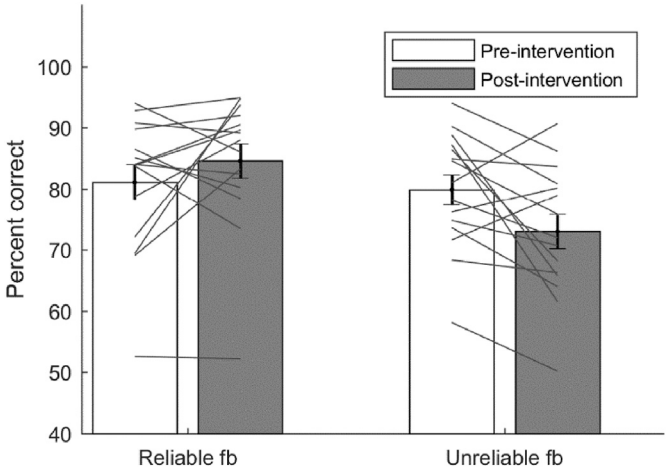


Fig. 2. Behavioural performance across groups (reliable/unreliable feedback) and test phases (pre-/post-intervention). The bars show mean performances and errorbars show standard errors of the means. The individual lines show subject-wise estimates of task performance.

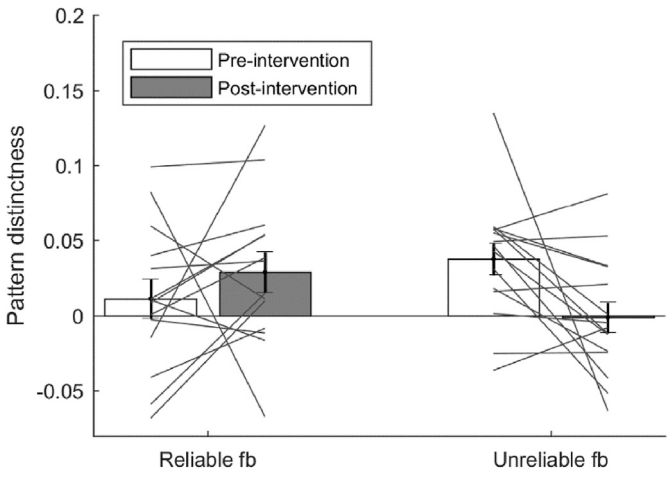


Fig. 3. Neural pattern distinctness in V1 plotted across groups (reliable/unreliable feedback) and test phases (pre-/post-intervention). The bars show the mean pattern distinctness and errorbars show standard errors of the means. The individual lines show subject-wise estimates of pattern distinctness.

significant decrease in pattern distinctness after unreliable feedback ($M = -0.04$, $SE = 0.02$, $t(14) = -2.61$, $p = .02$), but not after reliable feedback ($M = 0.02$, $SE = 0.02$, $t(13) = 0.98$, $p = .35$). The effect size of the change in pattern distinctness between the two groups was 0.84.

Since the pattern distinctness at each test phase (pre-/post-intervention) was averaged between the two stimulus pairs (i.e., pair 1: CW/CCW deviations from the 45° diagonal and pair 2: CW/CCW deviations from the 135° diagonal, see Fig. 1b), we tested for differences in the observed changes (post–pre) in pattern distinctness between them to investigate the possibility that the observed effects of feedback manipulation might have been driven by one stimulus pair alone. The comparison revealed that these pre- to post-intervention changes were comparable across the stimulus pairs for both the unreliable feedback group (paired *t*-test: $t(28) = 0.84$, $p = 0.41$; 45° reference: $M = -0.05$, SE = 0.01; 135° reference: $M = -0.03$, SE = 0.02) and the reliable feedback group (paired *t*-test: $t(26) = 1.21$, $p = 0.24$; 45° reference: $M = -0.003$, $SE = 0.03$; 135° reference: $M = 0.04$, $SE = 0.02$).

Next, to test whether the deterioration of neural pattern distinctness in V1 after unreliable feedback was related to the decrease in behavioural performance, we correlated the changes (post -intervention – pre-intervention) in performance with analogous changes in pattern distinctness in the unreliable feedback group (Fig. 4). The data were heteroscedastic (95% CI of differences between conditional variances from 600 iterations was [–93.65, −4.88]), making it possible that the observed correlations could be explained by the differences in variances. Hence, based on the recommendations of the Robust Correlation Toolbox (Pernet et al., 2013), we used a (1) a correlation estimate that is robust to multivariate outliers, i.e., the skipped Pearson's correlation and (2) a measure of significance that was robust to heteroscedasticity, i.e., the 95% CI obtained from bootstrapping the data (n = 1000).

This analysis revealed that there were no outliers and that the Pearson's correlation coefficient of 0.65 was associated with the 95% CI [0.31 0.86] which was well above zero and a p-value of .008 in the traditional Pearson correlation (in the absence of outliers, this analysis too yielded a correlation coefficient of 0.65) – indicating a clear positive correlation between the changes in performance and pattern distinctness following unreliable feedback interventions.

3.3. Unreliable feedback does not change overall neural activity

To understand if the observed changes in pattern distinctness is paralleled by a change in overall activity within the stimulus-responsive V1 voxels, the mean beta value was computed across all four stimulus types separately for the pre- and post-intervention phases of both groups. There was no detectable feedback-related change in the overall activity across these voxels, as shown by the absence of a significant interaction between feedback type and test phase ($F(1,26) = 0.26$, $p = .62$, $\eta_p^2 = 0.01$; Fig. 5). The main effects of feedback type ($F(1,26) = 0.85$, $p = .36$, $\eta_p^2 = 0.03$) and test phase ($F(1,26) = 0.39$, $p = .54$, $\eta_p^2 = 0.02$) did not reach significance. To further corroborate the absence of an interaction effect, we performed a Bayesian two-sample *t*-test comparing changes ($\Delta_{OverallActivity}$) between the two groups. The test showed anecdotal evidence for the absence of a group difference ($BF_{10} = 0.39$, $t(27) = .52$, $p = .61$, *95% CI* = [–35.54, 59.49]; default Cauchy prior provided by the package bayesFactor).

Further, we tested whether the overall activity could explain the changes in pattern distinctness reported in Section 3.2. by repeating the ANOVA with the difference in overall activity ($\Delta_{OverallActivity}$) as an additional covariate. The interaction between test phase and feedback type was still significant ($F(1,25) = 6.96$, $p = .01$, $\eta_p^2 = 0.22$). Further, the interaction between test phase (pre-/post-intervention) and the additional covariate was not significant ($F(1,25) = 2.22$, $p = .15$, $\eta_p^2 = 0.08$). Thus, the changes in neural pattern distinctness cannot be explained by

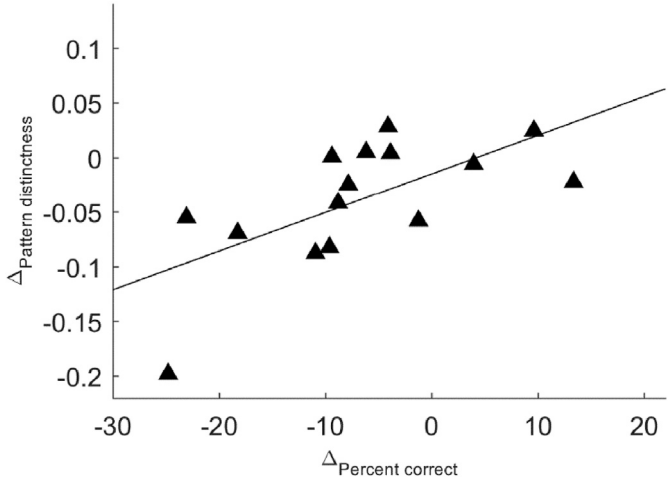


Fig. 4. The relationship between changes in behavioural performance ($\Delta_{Percent\ correct}$) and the precision of multivariate representations in V1 ($\Delta_{Pattern\ distinctness}$) for the unreliable feedback group. Triangles represent individual participants and the line shows the linear fit.

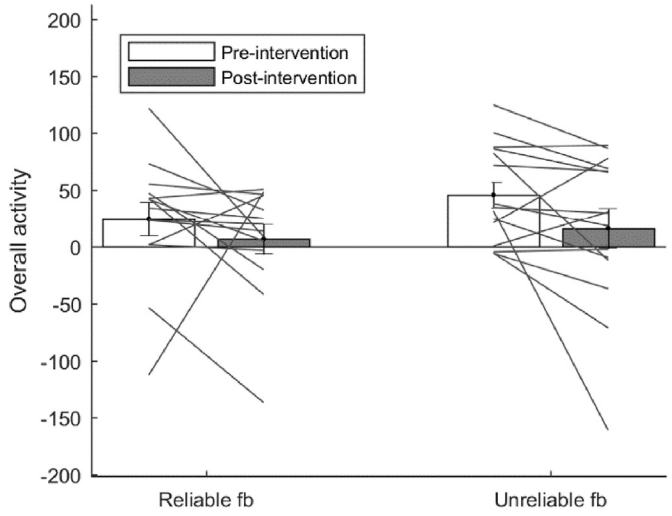


Fig. 5. Overall activity in stimulus-responsive V1 voxels plotted across groups (reliable/unreliable feedback) and test phases (pre-/post-intervention). The bars show the values averaged across participants, and errorbars show standard errors of the mean. Individual lines show subject-wise estimates.

changes in overall activity in the corresponding voxels.

3.4. Searchlight analysis shows a decrease in pattern distinctness in extrastriate visual areas

An exploratory whole-brain analysis was performed using the CV-MANOVA algorithm to yield searchlight maps for each test phase and participant. The changes in pattern distinctness in these maps (pre–post) were analysed at the group-level using a two-sample *t*-test. The results were then corrected for task-relevant voxels in the brain, and this revealed a cluster with a significant peak in the right visual association cortex (Fig. S1). No other significant peaks were detected. Details of this analysis are reported in the Supplementary Information.

3.5. Control analyses

We performed several control analyses to test whether non-perceptual factors might have influenced the observed behavioural and neural effects of unreliable feedback (see Supplementary Information). In brief, these control analyses yielded no evidence for a significant contribution of such non-perceptual factors. First, the analyses of the eye-tracking data revealed that fixation accuracy was comparable between pre-and post-intervention phases, and that the decodability of stimulus orientations (CW vs. CCW) from eye movements was at chance and did not change over time differently in the unreliable feedback group (Fig. S2). Next, the responses on the perceived reliability of feedback and the awareness of feedback manipulation were compared between groups. While there were group differences with respect to both the perceived reliability of feedback and the awareness of feedback manipulation, there was no detectable association between these ratings and the changes in performance and pattern distinctness in the unreliable feedback group (Fig. S3). Further, changes in motivation and the performance in the colour-change task were also unlikely to have been different between the two feedback groups (Figs. S4 and S5), making it highly unlikely that the feedback-related performance deterioration in the main task was due to a general decrease in motivation to do any task. Finally, we repeated the mixed ANOVA analyses of the performance and pattern distinctness data (described in Sections 3.1 and 3.2) with these potential confounds as additional covariates. Inclusion of eye-movement data, feedback awareness, or motivation rating as covariates did not abolish the significant interaction between feedback type and test phase for the behavioural or the neural data. Solely the inclusion of feedback reliability ratings rendered the interaction effect for performance insignificant, while the interaction effect for neural pattern distinctness remained significant. These results further show that the behavioural and neural effects of unreliable feedback were unlikely to be due to non-perceptual factors. More details on the control analyses are described in the Supplementary Information sub-section “Control analyses”.

4. Discussion

The current study used fMRI to investigate how unreliable feedback on perceptual task performance affects neural stimulus processing in early visual cortex. We predicted that the delivery of random unreliable feedback would be associated with a deterioration of perceptual performance, as previously shown (Herzog and Fahle, 1997; Vuvan et al., 2018; Varrier et al., 2019), and that this effect might be paralleled by a degradation of stimulus representations in V1. In line with our hypotheses, we observed that following the delivery of unreliable feedback, both task performance and neural pattern distinctness in V1 deteriorated, compared to a control group that received reliable feedback. Moreover, the changes in performance and pattern representation in the unreliable feedback group correlated with each other.

Together, these changes at the behavioural and neural level are thus comparable to a scenario in which stimulus uncertainty itself was manipulated, even though stimuli were kept constant throughout the experiment. This implies a mechanism by which the brain modulates sensory representations based on stimulus-independent factors. Indeed, such a mechanism is well-established for the induction of prior expectations about stimuli, including changes of neural stimulus representations in V1 (Kok et al., 2012, 2013). From a Bayesian perspective, this is compatible with an observer that integrates prior information and current sensory evidence, and weights each source of information based on its estimated reliability (precision). If expectations are strong or sensory information is deemed unreliable, sensory information is down-weighted.

Here we suggest that providing unreliable feedback to perceptual choices changes the observer’s belief about the reliability of sensory information, or hyperpriors (Friston et al., 2013), and thus leads to a down-weighting of sensory signals. This account would be well in line with our current and previous findings: (1) down-weighted sensory signal would lead to impaired task performance and reduced subjective confidence (Varrier et al., 2019) since the (task-relevant) sensory signals arriving at decisional and metacognitive stages are more noisy; (2) if feedback indeed affects the relative precision-weighting of sensory signals and prior information, a shift towards prior information would occur for unreliable feedback, as reported in Varrier et al., (2019); and (3) this relative shift away from sensory information would show a similar neural signature as reported for induced stimulus expectations. Yet, while we consider this a plausible interpretation of the observed behavioural and neural effects, it should be noted, that unlike our previous work (Varrier et al., 2019), the current study did not induce prior beliefs. We thus do not provide direct evidence for a shift of perceptual inference away from sensory evidence and towards prior beliefs.

As an alternative account, it is also conceivable that unreliable feedback might have reduced overall attention to the visual stimuli or motivation to perform the task, which could also explain the lower precision of stimulus representations in early visual cortex. However, the analysis of the mean activity, an indicator of attention (Kastner et al., 1998, 1999), computed across stimulus-responsive V1 voxels, showed that the pre-to post-intervention changes in performance and neural pattern distinctness were not related to overall neural stimulus responses. This indicates that our results are unlikely to be due to changes in attention. Similarly, further control analyses rendered confounding effects from changes in eye movements, subjectively experienced feedback reliability, awareness of the feedback manipulation and motivation unlikely. Thus, together with the correlation observed between the changes in performance and pattern distinctness in the unreliable feedback group, the decrease in the precision of stimulus representations at an early stage of cortical processing remains as a likely explanation of the decline in task performance after unreliable feedback. However, we also note that while the mixed ANOVAs revealed no significant effects of potential confounds, most of the Bayesian analyses showed only anecdotal evidence for the null hypothesis.

In studying the effects of unreliable feedback, our key assumption was that observers would attempt to infer and resolve the causes of the mismatch between their perceptual reports and external feedback. Hence it is indeed possible that participants could have also attributed these mismatches to a decrease in the reliability of response cues or feedback cues. However, by not informing participants of the response mapping (i.e., which button to press for which percept) until after the stimulus disappeared, we eliminate the role of response uncertainty in our measurement of representational precision (pattern distinctness) during stimulus presentation. Likewise, since feedback was absent altogether in the pre- and post-intervention test phases (where pattern distinctness was measured), this could not have influenced neural representations during stimulus presentation in the test phases either. In line with this, our results demonstrated that sensory processing was affected, as indicated by both the decrease in pattern distinctness in V1 and its correlation with corresponding changes in performance. The exploratory searchlight analysis also showed that stimulus representations in visual but not high-level executive brain areas changed as a result of unreliable feedback.

Nevertheless, in principle, there could still be a component of decision uncertainty that could have influenced task performance (i.e., "should I really press the button for CCW-tilt, when I saw CCW-tilt?") that may have gone undetected in the current experiment. Future studies can eliminate the decision noise component during stimulus presentation by, for instance, randomising the cognitive task itself across trials.

In the past, studies have shown that sampling efficiency improved with training (Lu and Dosher, 2004; Kurki and Eckstein, 2014; Moerel et al., 2016), and, since unreliable feedback has been associated with a deterioration in task performance (Herzog and Fahle, 1997, 1999; Varrier et al., 2019), it is possible that unreliable feedback leads to a decrease in sampling efficiency. However, unlike in the current study, these previous experiments (Lu and Dosher, 2004; Kurki and Eckstein, 2014; Moerel et al., 2016) trained over several sessions and used noise-overlays as a stimulus manipulation, and therefore proposed changes in stimulus sub-sampling as a mechanism for changes in perceptual learning. While we therefore consider a change in sampling efficiency less likely as a mechanism in in our study, such an interpretation would not contradict the notion of 'sensory down-weighting': A decrease in sampling efficiency for relevant sensory features would in fact be entirely in line with a mechanism by which the brain down-weights sensory channels that are inconsistent with the feedback information.

An alternative explanation of our results could also be that unreliable feedback hinders within-block learning that is often associated with such tasks (Herzog and Fahle, 1997; Liu et al., 2010). While this is possible in theory, we currently have no evidence to support this conjecture, since there was no significant improvement in performance or stimulus representations even in the reliable feedback group.

The exploratory whole-brain analysis revealed no change in stimulus representations in non-visual brain areas, but showed a decrease in the right extrastriate visual region. While this is different from our main ROI analysis of V1, it should be kept in mind that in the ROI analysis, pattern distinctness was determined based on voxels selected at the individual level. In contrast, the searchlight analysis used a single searchlight radius for all participants and was based on anatomically normalised functional images. These methodological differences make the searchlight analysis far less sensitive to the changes in the retinotopic stimulus representation in V1. Importantly, the searchlight analysis did not show significant changes in higher level brain areas (that were not sensitive to the physical properties of stimuli) as a result of unreliable feedback relative to reliable feedback.

One of the limitations of the current study is the participants' awareness of the feedback manipulation. However, control analyses showed that neither behavioural nor neural effects of unreliable feedback correlated with the awareness of the feedback manipulation. Higher awareness could have led to two processes and we do not find evidence for either of them: First, such awareness would have led to an overall reduced task motivation and attention and thereby decreasing task performance. However, neither the mean activity in stimulus-responsive V1 voxels, an indicator of top-down attention (Kastner et al., 1998, 1999), nor the subjective ratings of motivation showed evidence supporting a modulation of feedback type. Second, participants could have started ignoring feedback altogether when they noticed the feedback manipulation, and as a result performance would not suffer. In this case, participants who indicated higher awareness of feedback manipulation would have a smaller performance deterioration than those who indicated lower awareness of the manipulation. Yet, the control analysis indicated that there was no correlation between the ratings of awareness and the behavioural and neural changes in the unreliable feedback group. Also, subjective reports on the awareness of feedback manipulation provides only an imperfect control, since they are susceptible to participants' reporting bias – such as to indicate a higher awareness of feedback manipulation on being specifically asked about it during debriefing due to an "Aha!" effect. Hence, we recommend that future studies ensure better masking of the feedback manipulation, for instance by making the task more difficult.

A second limitation is the sample size used in the current study, given the between-group design. Since this study was novel in its design and approach, a formal sample size calculation could not be performed, and as a result, there is a risk of overestimation of the observed effect. The performance changes are comparable to our previous behavioural study using unreliable feedback (Varrier et al., 2019), which had a similar experimental design but with a within-subject design and a few other methodological differences. However, considering the novelty of the study design, the current results should be treated as proof-of-concept. Replication with larger sample sizes is therefore warranted.

According to predictive coding theories of hierarchical cortical processing, bottom-up sensory signals transmitted from lower to higher levels of the processing hierarchy are weighted by the precision of these signals in superficial cortical layers (Adams et al., 2013). While the precision, encoded by the post-synaptic gain of the neurons transmitting the bottom-up signals, is thought to be under the influence of top-down projections (Ryota et al., 2015), recent studies have shown evidence that these precision-weighted prediction errors could be encoded in sensory areas of the brain (Iglesias et al., 2013; Stefanics et al., 2019). We suggest that our finding of reduced distinctness of fMRI signal patterns may be due to a decrease in the precision-weighting of sensory information in V1, most likely mediated by top-down signalling of learned beliefs regarding the reliability of the sensory information.

5. Conclusion

The present study showed that the delivery of unreliable feedback resulted in impaired task performance and stimulus representations in early sensory areas of the brain. These changes could not be explained by changes in attention, motivation or eye movements. Awareness of feedback manipulation could have influenced our results although a detectable effect was not present; however, future studies should be cautious about this potential confound. Together with previous work, these results suggest that unreliable feedback can change the beliefs about the uncertainty of sensory information, entailing behavioural and neural effects that are highly similar to those reported for increases of stimulus-based uncertainty.

CRediT authorship contribution statement

Rekha S. Varrier: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Funding acquisition. **Marcus Rothkirch:** Resources, Writing - review & editing. **Heiner Stuke:** Conceptualization, Writing - review & editing. **Matthias Guggenmos:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Philipp Sterzer:** Conceptualization, Methodology, Funding acquisition, Supervision, Project administration, Writing - review & editing.

Acknowledgements

This study was supported by the German Research Foundation (DFG) grants GRK 1589/2, STE 1430/6-2, STE 1430/8-1 and GU 1845/1-1. Heiner Stuke is participant in the Charité Clinician Scientist Program funded by the Charité – Universitätsmedizin Berlin and the Berlin Institute of Health. We also thank Musaab Assel and Katharina Kanthak for their help with data acquisition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116701.

References

Adams, R.A., Stephan, K.E., Brown, H.R., Frith, C.D., Friston, K.J., 2013. The computational anatomy of psychosis. Front. Psychiatr. 4. Accessed August 23, 2018. https://www.frontiersin.org/articles/10.3389/fpsyt.2013.00047/full.

Allefeld, C., Haynes, J.-D., 2014. Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. Neuroimage 89, 345–357.

Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E., Malach, R., 2002. Contrast sensitivity in human visual areas and its relationship to object recognition. J. Neurophysiol. 87, 3102–3116.

Boynton, G.M., Demb, J.B., Glover, G.H., Heeger, D.J., 1999. Neuronal basis of contrast discrimination. Vis. Res. 39, 257–269.

Darcy, N., Sterzer, P., Hesselmann, G., 2019. Category-selective processing in the two visual pathways as a function of stimulus degradation by noise. Neuroimage 188, 785–793.

Durlak, J.A., 2009. How to select, calculate, and interpret effect sizes. J. Pediatr. Psychol. 34, 917–928.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25, 1325–1335.

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 815–836.

Friston, K.J., Lawson, R., Frith, C.D., 2013. On hyperpriors and hypopriors: comment on Pellicano and Burr. Trends Cognit. Sci. 17, 1.

Gardumi, A., Ivanov, D., Hausfeld, L., Valente, G., Formisano, E., Uludağ, K., 2016. The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. Neuroimage 132, 32–42.

Haynes, J.-D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat. Neurosci. 8, 686–691.

Hebart, M.N., Donner, T.H., Haynes, J.-D., 2012. Human visual and parietal cortex encode visual choices independent of motor plans. Neuroimage 63, 1393–1403.

Hendriks, M.H.A., Daniels, N., Pegado, F., Op de Beeck, H.P., 2017. The effect of spatial smoothing on representational similarity in a simple motor paradigm. Front. Neurol. 8, 222.

Herzog, M.H., Fahle, M., 1997. The role of feedback in learning a vernier discrimination task. Vis. Res. 37, 2133–2141.

Herzog, M.H., Fahle, M., 1999. Effects of biased feedback on learning and deciding in a vernier discrimination task. Vis. Res. 39, 4232–4243.

Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., den Ouden, H.E.M., Stephan, K.E., 2013. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron 80, 519–530.

Kahnt, T., Grueschow, M., Speck, O., Haynes, J.-D., 2011. Perceptual learning and decision-making in human medial frontal cortex. Neuron 70, 549–559.

Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685.

Kastner, S., Weerd, P.D., Desimone, R., Ungerleider, L.G., 1998. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science 282, 108–111.

Kastner, S., Pinsk, M.A., De Weerd, P., Desimone, R., Ungerleider, L.G., 1999. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22, 751–761.

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

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

Kok, P., Brouwer, G.J., van Gerven, M.A.J., de Lange, F.P., 2013. Prior expectations bias sensory representations in visual cortex. J. Neurosci. 33, 16275–16284.

Krekelberg, B., 2019. klabhub/bayesFactor. KLab GitHub. Accessed January 11, 2020. https://github.com/klabhub/bayesFactor.

Kurki, I., Eckstein, M.P., 2014. Template changes with perceptual learning are driven by feature informativeness. J. Vis. 14. Accessed January 28, 2020. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4160004/.

Liu, J., Lu, Z.-L., Dosher, B.A., 2010. Augmented Hebbian reweighting: interactions between feedback and training accuracy in perceptual learning. J. Vis. 10, 29–29.

Lu, Z.-L., Dosher, B.A., 2004. Perceptual learning retunes the perceptual template in foveal orientation identification. J. Vis. 4, 5–5.

Ludwig, K., Sterzer, P., Kathmann, N., Hesselmann, G., 2016. Differential modulation of visual object processing in dorsal and ventral stream by stimulus visibility. Cortex 83, 113–123.

Mamassian, P., 2016. Visual confidence. Annu. Rev. Vis. Sci. 2, 459–481.

Misaki, M., Luh, W.-M., Bandettini, P.A., 2013. The effect of spatial smoothing on fMRI decoding of columnar-level organization with linear support vector machine. J. Neurosci. Methods 212, 355–361.

Moerel, D., Ling, S., Jehee, J.F.M., 2016. Perceptual learning increases orientation sampling efficiency. J. Vis. 16, 36–36.

Op de Beeck, H.P., 2010. Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? Neuroimage 49, 1943–1948.

O'Reilly, J.X., Jbabdi, S., Behrens, T.E.J., 2012. How can a Bayesian approach inform neuroscience? Eur. J. Neurosci. 35, 1169–1179.

Pernet, C.R., Wilcox, R.R., Rousselet, G.A., 2013. Robust correlation analyses: false positive and power validation using a new open source Matlab toolbox. Front. Psychol. 3. Accessed September 29, 2019. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00606/full#h6.

Ryota, Kanai, Yutaka, Komura, Stewart, Shipp, Karl, Friston, 2015. Cerebral hierarchies: predictive processing, precision and the pulvinar. Phil. Trans. Roy. Soc. B Biol. Sci. 370, 20140169.

Series, P., Seitz, A., 2013. Learning what to expect (in visual perception). Front. Hum. Neurosci. 7. Accessed January 30, 2020. https://www.frontiersin.org/articles/10.3389/fnhum.2013.00668/full.

Stefanics, G., Stephan, K.E., Heinzle, J., 2019. Feature-specific prediction errors for visual mismatch. Neuroimage 196, 142–151.

Sterzer, P., Frith, C., Petrovic, P., 2008. Believing is seeing: expectations alter visual awareness. Curr. Biol. 18, R697–R698.

Summerfield, C., Monti, J.M.P., Trittschuh, E.H., Mesulam, M.-M., Egner, T., 2008. Neural repetition suppression reflects fulfilled perceptual expectations. Nat. Neurosci. 11, 1004–1006.

Thielen, J., Bosch, SE, Leeuwen, TM van, Gerven, MAJ van, Lier, R van, 2019. Evidence for confounding eye movements under attempted fixation and active viewing in cognitive neuroscience. Sci. Rep. 9, 1–8.

Tong, F., Harrison, S.A., Dewey, J.A., Kamitani, Y., 2012. Relationship between BOLD amplitude and pattern classification of orientation-selective activity in the human visual cortex. Neuroimage 63, 1212–1222.

Varrier, R.S., Stuke, H., Guggenmos, M., Sterzer, P., 2019. Sustained effects of corrupted feedback on perceptual inference. Sci. Rep. 9, 5537.

Vuvan, D.T., Zendel, B.R., Peretz, I., 2018. Random feedback makes listeners tone-deaf. Sci. Rep. 8. Accessed January 28, 2020. http://www.nature.com/articles/s41598-018-25518-1.