No effect of feedback, level of processing or stimulus presentation protocol on perceptual learning when easy and difficult trials are interleaved.

Jordi M Asher[1], Paul B Hibbard

Department of Psychology, University of Essex

Abstract

The role of feedback during training is a topic of great theoretical importance in perceptual learning. Feedback can be provided externally by the environment or internally by the observer. In order to evaluate the effectiveness of learning with internal versus external feedback, we performed a large multi-level experiment, varying the type of training task (Motion or Form), the level of processing (Local or Global), the presence of feedback (With or Without) and finally the method of stimulus presentation (Adaptive staircase or Method of constant stimuli). 140 participants were randomly assigned to one of ten groups and undertook 3 days of training in one condition only. Detection thresholds were measured daily before and after training with a pre- and post-assessment. A 75% detection threshold was calculated and used to estimate that day's training levels (65% and 85% accuracy for difficult and easy trials respectively). The group trained with MOCS were presented with predefined randomly interleaved easy and difficult trials ranging from 50% to 95% stimulus intensity. Our findings indicate that improvement was generally robust across training-tasks, processing levels and feedback conditions. This suggests that internal reinforcement is as effective as external feedback in a discrete-noise-paradigm for local and global tasks when easy and difficult trials are interleaved.

Keywords: perceptual learning, local motion, local form, global motion, global form, psychophysics, external feedback, internal feedback, equivalent noise.

Introduction

The role of feedback in perceptual learning is a topic of rich theoretical and empirical importance, in understanding the scope and mechanisms of neuroplasticity (Seitz et al., 2006; Shibata et al., 2009; Herzog et al., 2006; Dosher & Lu, 1999; Herzog & Fahle, 1997). Training-induced improvement in perception has been found across all perceptual modalities,

*Corresponding author
Email address: jashera@essex.ac.uk (Jordi M Asher)

including sight (Karni & Sagi, 1991), touch (Dinse et al., 2003), taste, smell (Wilson & Stevenson, 2003; Chu et al., 2016) and hearing (Harrison et al., 2005). These improvements can have important benefits in fields including radiography (Kundel et al., 2007), optometry (Sterkin et al., 2018), sports science (Deveau et al., 2014) and education (Kellman, 2013; Ahissar, 2007).

The rate and efficiency of perceptual learning can be improved if external feedback is provided to the observer (Herzog & Fahle, 1997; Garcia et al., 2013; Dobres & Watanabe, 2012; Herzog & Fahle, 1998). Feedback refers to any signal that provides the observer with information about the quality of their behaviour or response towards a target (Dosher & Lu, 2017). Typically in studies of perceptual learning, feedback is provided in forced choice experiments via an auditory or visual cue to inform the observer as to the accuracy of their response. However, perceptual improvements have also been reported in learning experiments in the absence of external feedback (Ball & Sekuler, 1982; Fahle, 2005; Liu et al., 2010; Petrov et al., 2006; Vaina et al., 1998; Shiu & Pashler, 1992). Furthermore, learning does not always occur, even after extensive training with explicit feedback (Cumming et al., 1998). External feedback is therefore neither necessary nor sufficient for learning to occur. Nevertheless, establishing its role in learning is important in understanding the mechanisms of perceptual learning, and in devising efficient and effective training strategies.

Despite this, the role of feedback is typically not of primary interest in perceptual learning studies, so its role is still to a large degree unclear. Only a handful of studies have explicitly manipulated the use of feedback in learning, and these have produced some conflicting and equivocal results regarding the necessity of feedback (Seitz et al., 2006; Liu et al., 2012; Asher et al., 2019; Herzog & Fahle, 1997; Shibata et al., 2009; Liu et al., 2014; Aberg & Herzog, 2012). These studies are difficult to compare, owing to a variety of methodological differences employed, such as the psychophysical methods and corresponding measures of learning; the methods used to deliver feedback; the level of difficulty; and the nature and complexity of the perceptual task.

Psychophysical procedures

Multiple psychophysical procedures, and criteria for demonstrating successful learning, have been employed in these studies. Some vary the stimulus level over time so as to ensure a constant level of performance, and track the resulting change in the stimuli needed to maintain this performance level (Petrov et al., 2005, 2006). Other studies keep the stimulus level constant and measure how performance changes over time (Vaina et al., 1995; Herzog & Fahle, 1997), or present a fixed range of stimuli and measure changes in threshold (the stimulus needed to ensure a criterion level of performance) (Liu et al., 2012; Karni & Sagi, 1991; Ahissar & Hochstein, 1997). Typically, the effect of feedback is strongest when provided after each trial. However, providing an indication of the proportion of correct trials at the end of a block can also improve learning, albeit to a lesser extent (Shibata et al., 2009; Aberg & Herzog, 2012). Herzog & Fahle (1997) evaluated the role of feedback, using a vernier acuity task by training six groups of participants around threshold. Learning improved substantially with feedback. There were large individual differences in performance for those who trained without feedback; some showed a small improvement, others did not improve, got worse or oscillated. This has raised interesting discussions on persistent and transient learning (Aberg & Herzog, 2009; Dosher & Lu, 2017). Learning has also been found in the absence of external feedback (Ball & Sekuler, 1982; Fahle, 2005; Liu et al., 2012; Petrov et al., 2006; Vaina et al., 1998; Shiu & Pashler, 1992), specifically when the task is easy, but not when only difficult trials are presented (Liu et al., 2010). While learning without feedback has been found when easy and difficult trials are interleaved (Fahle & Edelman, 1993; Liu et al., 2012), this is not always the case (Seitz et al., 2006; Asher et al., 2019).

Studies have also varied the stimuli and task used in training. Liu et al. (2012) found perceptual learning without feedback on a task requiring the discrimination of the orientation of a noised-masked target. Seitz et al. (2006) studied the role of feedback using two tasks. In the first, observers reported the direction of motion. Task difficulty was manipulated by varying the contrast of the stimulus, with 10 contrast levels presented. Observers showed significant improvement after one day of training when trial-by-trial feedback was provided, but not when there was no feedback. In the second task, observers reported the orientation of a bar stimulus, and task difficulty was manipulated by the addition of a random dot mask, with 8 levels of signal-to-noise ratio. After 10 days of training, they again found improvement in performance only if feedback was provided. Liu et al. (2012) found perceptual learning without feedback, provided that the training contained the appropriate combination of task difficulty levels. They used a similar task to Seitz et al. (2006), in which observers judged the orientation of a target that was masked by noise. However, rather than presenting fixed levels of signal-to-noise ratio, they used an adaptive staircase technique to maintain the performance level at either 65% or 85% accuracy. They found an improvement in performance at both stimulus levels after training without feedback, but only if both easy (85% accuracy) and difficult (65% accuracy) trials were included in the training.

In a typical perceptual learning experiment, such as those described above, the observer performs a forced-choice discrimination task, and learning is measured as an increase in accuracy. This improvement can be understood as the result of an updating of the weights connecting the encoded stimulus, via a multi-level neural network of visual processing and categorisation (Petrov et al., 2005; Dosher & Lu, 2009). These weights can be updated to reinforce those connections that lead to successful classification, and to diminish those that do not. When external, trial-by-trial feedback is provided, this may be used to guide the learning process. In the absence of external feedback, it has been proposed that easy trials can act to 'bootstrap' the more difficult trials (Fahle & Edelman, 1993). Petrov et al. (2005) proposed that, in the absence of external feedback, weights are updated under the assumption that the response is accurate. If performance is sufficiently accurate prior to training, this will result in perceptual learning.

Learning with easy and difficult exemplars

A wealth of perceptual learning studies support the idea that training on easy-to-discriminate trials facilitates learning for difficult trials (Ahissar & Hochstein, 1997; Liu et al., 2010, 2012; Liu & Weinshall, 2000; Garcia et al., 2013). In an orientation discrimination task, Ahissar & Hochstein (2004) manipulated task difficulty by decreasing the stimulus processing time available between the presentation of a target and a distractor, and found that, as difficulty increased, improvement was slower and specific to the trained orientation. However, in the conditions with increased time between stimuli and distractors, which increased the number of correct responses, learning was quicker and transferred across orientations. Training was also more effective when there was a gradual transition from easy to difficult trials (Ahissar & Hochstein, 2004). Investigating the time course for learning to detect global form in Glass patterns, Garcia et al. (2013) found that, consistent with prior evidence, training on easy tasks facilitated performance on later difficult trials. They also found that performance improvements were greater with those observers trained on difficult tasks, even when those trained with easy tasks received twice as much training (Garcia et al., 2013). Similarly, Liu et al. (2012) found that when training on an orientation discrimination task that was set at a low accuracy (65% correct), learning only occurred when external feedback was provided. In contrast, when high accuracy (85% correct) trials were used, performance improved whether feedback was present or not. When high and low accuracy trials were mixed, performance improved for both of the accuracy levels, for both the feedback and the no-feedback conditions (Liu et al., 2012). In contrast to studies that find improvement after gradually transitioning from easy to difficult trials (Ahissar & Hochstein, 2004; Garcia et al., 2013), difficult and easy trials were randomly interleaved. This suggests that the presence of easy trials may be sufficient to provide internal feedback, and bootstrap more difficult trials.

Talluri et al. (2015) proposed a reweighting model of perceptual learning in which the degree of reweighting is moderated by the confidence in the perceptual decision on each trial. This allows greater reweighting when confidence is higher, and the response thus more likely to be correct. In this model, confidence is quantified as the strength of the output response of the network. Alternatively, it is possible that explicit meta-cognitive signals of perceptual confidence (Mamassian, 2016) are used to guide reweighting. It is known that choice-based decisions may be influenced by internal body states via interoceptive information (Gu & FitzGerald, 2014), which can act as an internal sense of subjective confidence (Herzog & Fahle, 1997; Petrov et al., 2005). A recent fMRI study observed that brain areas typically associated with reward signals from externally generated feedback signals, were also activated when confidence was high (Guggenmos et al., 2016). Confidence based perceptual learning may occur as a result of the brain reinforcing those behaviours that produce states of high confidence, and reducing those that result in low confidence (Guggenmos et al., 2016; Ott et al., 2018). This evidence is in line with computational models that suggest a relationship between learning as a result of decision confidence (Talluri et al., 2015) and through increased probabilistic inference (Bejjanki et al., 2011).

The influence of meta-cognitive confidence signals could potentially explain why learning does not always occur when external feedback is provided. Whereas Liu et al. (2012) used just two stimulus levels during training and found perceptual learning without feedback, Seitz et al. (2006) used multiple levels, and only found learning when feedback was provided. Similarly, Asher et al. (2019) assessed the role of external feedback in a global motion task. Stimuli were presented at multiple levels, using the method of constant stimuli, and learning only occurred when feedback was provided. Although the overall performance level of 75% was the same as that of Liu et al. (2012), one key difference is that Asher et al. (2019) presented stimuli at 7 levels, while Liu et al. (2012) used just two. This difference may be important since confidence ratings are only accurate when stimulus levels are easily distinguishable. When presented with multiple stimulus levels, observers tend to be overconfident in difficult trials, and underconfident in easy trials (Zylberberg et al., 2014). This predicted inaccuracy in confidence judgements would reduce the efficiency of learning, if it is taken into account during reweighting (Talluri et al., 2015). This would then account for the robust learning found when only two stimulus levels were presented during training (Liu et al., 2012), and the lack of learning found with the multiple stimulus levels used in the method of constant stimuli (Seitz et al., 2006; Asher et al., 2019).

Level of processing

Another important consideration when understanding the role of feedback in perceptual learning is the nature of the task and level of processing required. Visual processing is understood to be functionally specialised, hierarchically organised and connected by feedforward, feedback and lateral connections between areas (Zeki, 2003; Newsome et al., 1989; Maunsell et al., 1990; Britten et al., 1993; Rudolph & Pasternak, 1999; Felleman & Van Essen, 1991; Lennie, 1998; Livingstone & Hubel, 1988; Callaway, 2004). The receptive fields of neurons become more functionally specific the higher up the hierarchy they are (Zeki, 1978), and their size and function differ based on their location and specialisation. For example, the receptive fields in the lowest cortical areas, V1 and V2, represent simple visual dimensions, such as the position, orientation and scale of local image features (Zeki, 2003; Lennie, 1998; Gilbert et al., 2001), over a small area of the visual field (Furlan & Smith, 2016; Lamme, 2003; Simoncelli & Heeger, 1998). By contrast, receptive fields in higher cortical areas such as V3, V4 and V5, are much larger, and pool the incoming sensory information from lower level receptive fields (Burr & Thompson, 2011; Gilbert et al., 2001) across features such as space, time, speed, direction of motion and spatial frequency (Burr & Thompson, 2011; Gilbert et al., 2001; Felleman & Van Essen, 1987; Furlan & Smith, 2016; Hubel & Wiesel, 1965; Mikami et al., 1986; Movshon et al., 1978; Sillito et al., 2006; Zeki, 1974).

Local and global tasks have been developed in order to selectively assess these levels of processing (Newsome et al., 1989; Braddick, 1993; Gilbert et al., 2001; Bex & Dakin, 2002; Ostwald et al., 2008; Amano et al., 2009; Burr & Thompson, 2011; Nishida, 2011). It should be noted that formal definitions of tasks for "local" and "global" processing do not exist, and the distinction between the two is not always clear cut, and therefore we propose these working definitions of local and global processing tasks. A *local* processing task is comprised of a signal that is congruent across all dimensions, and could, in principle, be performed using a single feature of the stimulus. In contrast, the stimulus elements in a *global* processing task is comprised of a signal that is incongruent (e.g. noisy) across its dimensions (i.e. spatial and/or temporal). In this situation the task could *not* be reliably performed using a single feature of the stimulus (see figure 1). Neural responses to a local task will be highly congruent across stimulus elements, a global task will have lower congruency (e.g. containing noise) of stimulus elements across spatial locations.

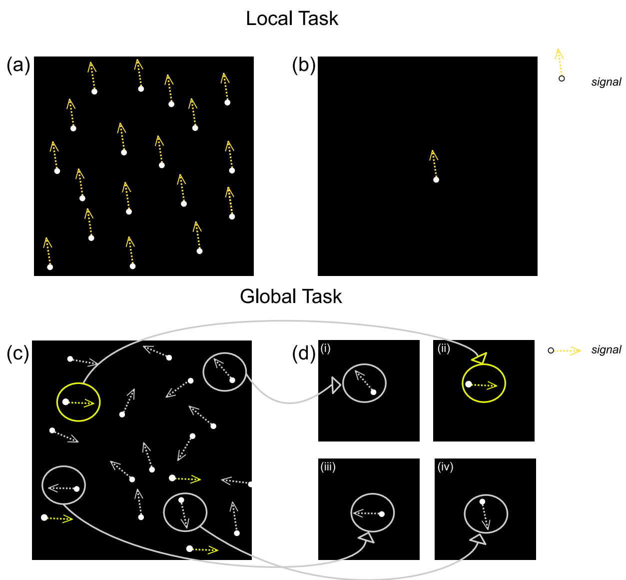

Figure 1: Typical motion tasks, where a number of dots (white) are presented on screen, observers are required to integrate the signal across all spatial locations and determine the general direction (left/right) of motion (arrows indicate the trajectory). (a) A local processing motion task comprised of up-leftward motion from the vertical mid-point. All the stimulus elements move coherently across spatial locations. (b) When only supplying one (randomised) dot direction discrimination is not impeded. (c) A global processing motion task, comprised of a signal (congruently moving elements) and noise (incongruently moving elements). An observer will need to integrate the conflicting signal across the spatial location (i.e. direction of motion). (d i-iv) Making a direction judgement from one randomised dot would not lead to reliable performance.

Stimuli choices for this set of experiments

In order for all of our experimental conditions to be as similar to each other as possible, we considered the motion and form tasks widely used in the literature. There is general consensus in what defines a motion task; i.e. elements moving in the same direction or varied directions. Global motion is typically assessed using random dot kinematograms (RDK) that implement one of two common strategies to create incoherent motion. The first strategy creates incoherent motion by initially setting the direction of motion of each individual dot as the signal direction. This is then perturbed by an additional noise direction, for example sampled from a Gaussian distribution of directions (Newsome et al., 1989). In this case, each dot carries both the signal and the noise. the signal direction is defined as the mean direction across the whole population of dots. Difficulty is increased by manipulating the standard deviation of the distribution of directions presented (Tibber et al., 2014). This type of task is often referred to as an *equivalent-noise paradigm* (Mansouri & Hess, 2006; Dakin et al., 2005; Tibber et al., 2014) and operates as a "population-vector-average" which determines a weighted average of the strength of the neural responses (Dakin et al., 2005). In a second variant, dots are randomly segmented into signal and noise categories. All signal dots move in the same direction, and the noise dots move in random directions. The observer's task is to determine the signal direction (Scase et al., 1996). A further complexity in this case is that, in addition to the integration of signal dots across the stimulus, these need also to be segmented from the noise dots (Braddick, 1993). Difficulty is increased by reducing the ratio of signal-to-noise dots and operates using a "winner-take-all" strategy, which identifies which directionally-tuned channel is the most active (Dakin et al., 2005). We define a comparable working definition for this as *discrete-noise paradigm*.

However, for shape, the definition is more complex. Tasks have been developed to understand how local orientation signals make a global percept of a shape or structure (Badcock et al., 2013; Chung & Khuu, 2014). In these tasks, a local signal is defined by orientation, either of small, oriented targets (Braddick et al., 2000), or by pairs of dots (diapoles) to generate a global structure (Ostwald et al., 2008). Glass patterns (Glass, 1969), for example, use random dots to generate a global form pattern (i.e. concentric, radial, hyperbolic etc) (Wilson & Wilkinson, 1998). Global form tasks may be complicated further if the signal is defined by a pattern in which the direction of motion or orientation is not the same in all parts of the stimulus (Graziano et al., 1994; Braddick et al., 2000). By varying this direction across the stimulus, the signal might for example define a spiral, or a diverging or converging motion pattern.

In order to create a directly comparable stimulus across tasks and processing levels, that also avoids the complexities of equivalent noise paradigms and shape patterns, we selected a discrete-noise paradigm that defines the ratio of signal to noise. Since our local task contains 100% signal we manipulate difficulty by reducing discriminability to chance performance (left/right decision).

The distinction between local and global tasks has been considered in the context of perceptual learning. Ahissar & Hochstein (2004) noted that, since the global stages of processing integrate over space and scale, they tend to have broader tuning properties. As such, any learning associated with the global stage of processing is predicted to transfer more broadly to stimuli with different characteristics. Asher et al. (2019) showed that learning for global motion exhibits the spatial frequency tuning of the global stage of processing. Levi et al. (2015) showed transfer of learning from a global motion task to other, unrelated tasks including contrast sensitivity, and Huxlin et al. (2009) showed perceptual relearning following training with global motion stimuli in cases of cortical blindness. Despite the potential for greater generalisation of learning when trained on global tasks, there is also evidence that greater plasticity exists in the early, local stages. Watanabe et al. (2002) found greater perceptual learning as a result of passive viewing for local than for global tasks. This reduced plasticity for global motion stages may reflect the greater complexity of processing, requiring the integration and segmentation of information, and therefore the more indirect link between the encoded visual stimulus and the decision making stage.

Summary

There are conflicting results from studies that report on perceptual learning when easy and difficult trials are interleaved without trial feedback (Liu et al., 2012) and those that do not (Asher et al., 2019; Seitz et al., 2006). However, owing to the variation in the methods it is difficult to directly compare the findings of these studies. In order to disentangle the conflicting results, we considered a number of factors that may account for the differences in the findings; including (i) the level of processing (local/global) and the type of task (motion/form), (ii) the provision of trial-by-trial feedback and (iii) the method of presentation of the stimuli (e.g. adaptive staircase or MOCS), and designed tasks to be as similar as possible.

Aims

The aim of the current study was to assess two factors which affect the ability of the visual system to learn in the absence of external feedback. The first factor is the distinction between local and global processing stages. Given the greater plasticity of local processing stages (Watanabe et al., 2002), we predict that external feedback will be more critical for perceptual learning for global than for local tasks.

The second factor is the nature of the psychophysical procedure used during training. While some studies have used just one or two stimuli during the training stage (e.g. Liu et al. (2012)), others have presented many levels, for example when using the method of constant stimuli (Asher et al., 2019; Seitz et al., 2006). While this might seem a minor methodological detail, it may have important implications for accurate meta-cognitive confidence judgements (Zylberberg et al., 2014), and thus the ability to update network weights appropriately on the basis of accurate responses (Talluri et al., 2015). These factors may be responsible for the inconsistent results found in previous studies (Seitz et al., 2006; Liu et al., 2012; Asher et al., 2019).

We performed three experiments to address these two questions. In the first experiment, we assessed the importance of feedback in perceptual learning for local and global form tasks, following training using a combination of easy and difficult trials. In the second experiment, this was repeated for local and global motion tasks. In our third experiment, we assessed perceptual learning for the same global motion task, but used the method of constant stimuli to present the training trials. We predicted that perceptual learning would be successful in all cases when external feedback was provided. When no feedback was provided, we expected learning to be reduced or absent (i) for global compared with local tasks, and (ii) when training used the method of constant stimuli.

General Methods

Participants

140 observers from the University of Essex participated in the study, in one of ten conditions, resulting in a total of 14 observers in each condition. All observers had normal or corrected-to-normal vision and were paid for their participation or received course credit. The procedures were approved by the University of Essex University Ethics Committee (JA1609). All observers gave informed written consent.

Apparatus

Stimuli were presented on a 27" iMac with a display resolution of 2560x1440 pixels and 60 Hz refresh rate. Stimuli were generated and presented with Matlab 2015a, using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) on a 2.7 Ghz IMac running OSX 10.9.5. Stimuli were viewed from a distance of 500 mm, and one pixel subtended 1.6'(arcminutes).

General Stimuli and Procedural information

Stimuli were 4.8' $\times$ 4.8' square white dots displayed within a black 7.95 $^{\circ}\times$ 7.95° square region in the centre of a black screen. For the motion conditions, there were 225 white dots and for the form tasks there were 225 white dot pairs (450 dots). Half of the observers were provided with trial-by-trial feedback during the training blocks. When feedback was present, this was provided as an immediate auditory beep after each trial, a high pitched tone for a correct response (2000Hz for 100ms), and a low pitched tone for an incorrect response (200 Hz for 400 ms). No feedback was provided for any group during the pre- and post-assessment phases.

The general procedure for all conditions was identical. Observers were trained for three consecutive days on one condition only, either with or without feedback. Each observer completed daily baseline measures before and after training for the same task, without feedback (see Figure 2). All observers were briefed orally, read written instructions on the task and finally completed a short demonstration. All testing was performed in a dark room and observers were regularly prompted to check their distance from the screen with a measured piece of string that was affixed to the monitor. A 30 second break was provided before and after the training block, and 10 second breaks were provided in between each of the 12 training blocks, although observers were advised that they could take longer breaks if required.

Each day of data collection consisted of (1) one block of trials to estimate the threshold followed by (2) 12 blocks of training trials and finally (3) one block of trials to estimate the threshold after training.

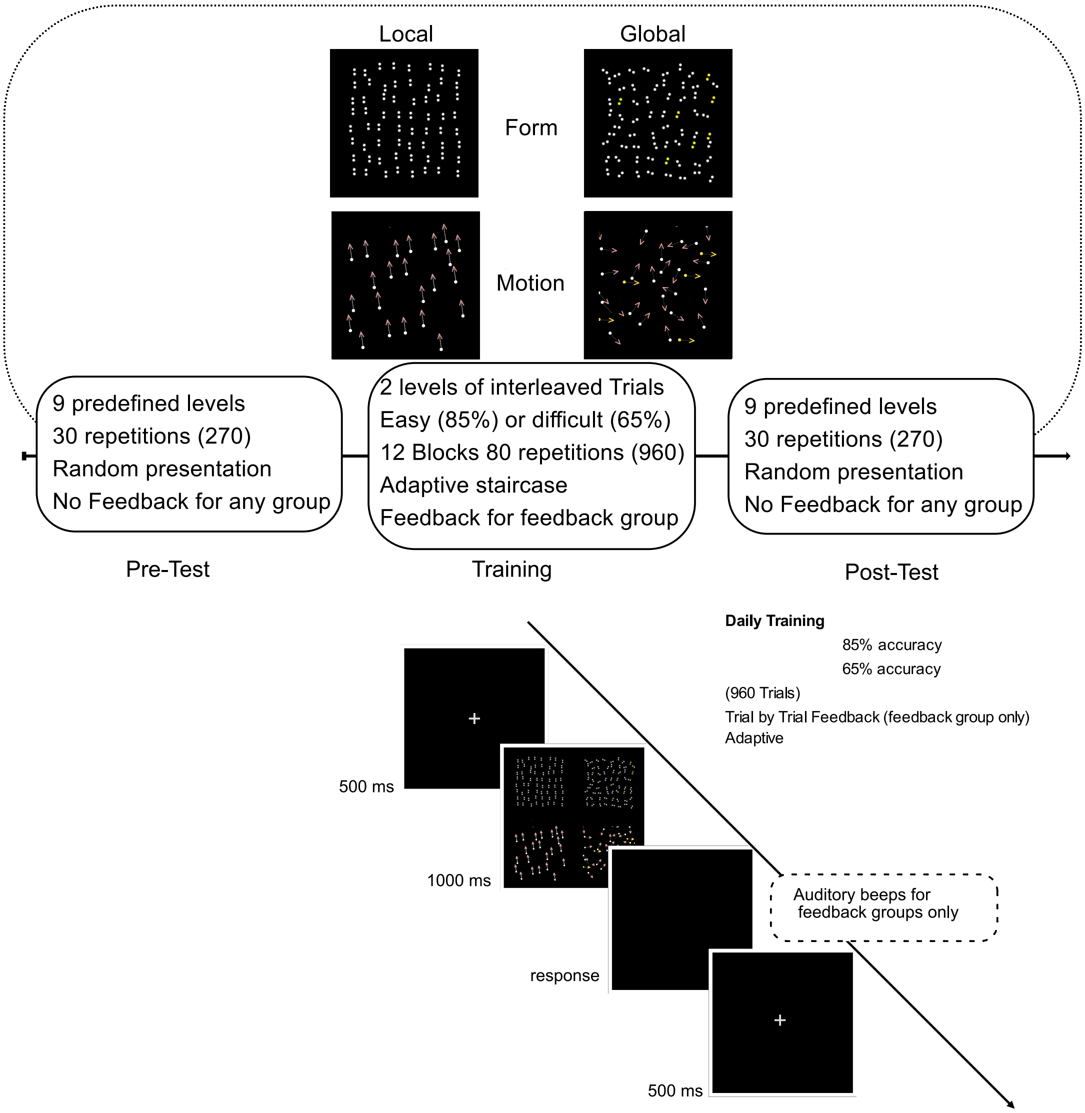

Figure 2: Psychophysical design and procedure. Larger examples of the stimuli appear later in the text. 140 observers were randomly assigned to one of ten groups and undertook three days of training in one condition only. All observers undertook a daily pre- and post-assessment. The baseline assessments were presented using MOCS at 7 predefined stimulus intensities. A 75% detection threshold was calculated and used to estimate that day's training levels (65% and 85% accuracy for difficult and easy trials respectively). The group trained with MOCS were presented predefined randomly interleaved easy and difficult trials ranging from 50% to 95% stimulus intensity. There were 12 training blocks with 80 trials each. Feedback was presented only during the training phase, and only to observers in a feedback-present group.

Experiment 1 - Local and Global Form

The individual dots in each pair were separated by 9.6' (see Figure 3 (a)). Form was defined by manipulating the direction from one dot to the other. In the local task, each dot pair had the same orientation. In the global task, a subset of signal dot pairs had the same orientation, while the remaining noise dot pairs were oriented randomly. So that dot pairs did not overlap, dots were rendered on an invisible grid at the centre of the screen. Each of the square regions was 0.53×0.53 ° and contained a single randomly positioned dot pair. Each stimulus was presented for 1 second, and the observers' two alternative forced choice task was to indicate whether the orientation of the stimulus was rotated clockwise or anti-clockwise relative to vertical. The next trial did not begin until a response was made. A fixation cross was then presented in the centre of the screen for 500 ms, followed by the next trial after a further 500ms pause.

In the pre- and post-training blocks, stimuli were presented using the method of constant stimuli (MOCS), and observers responded by pressing the left or right arrow on the keyboard to indicate the orientation of the stimulus. The order of trials was randomised, and contained 30 repetitions for each of nine stimulus levels (270 trials per block). Testing was performed before and after training each day. For the local task, dots were either oriented on the vertical (0°), or tilted at one of the following angles: ± 5°, 2.5°, 1.25°, 0.625°. For the global task, dot pairs were randomly designated as signal or noise. The number of signal dot pairs for each trial was randomly selected from the predefined levels (5, 10, 15, 37, 60, 75, 100, 125, 150 pairs). The remaining dot pairs were then designated as noise pairs for that trial. All signal dot pairs were orientated either to left or right of vertical at ± 20° for each trial. Each noise dot pair was randomly oriented between 0-360°. Example stimuli are shown in Figure 3 (b). No feedback as to the accuracy of responses was provided in the pre- and post-training blocks in any of the conditions.

Following the pre-training block, a cumulative Gaussian curve was fit to each observer's responses each day using the Palamedes toolbox (Prins & Kingdom, 2009) in order to determine 65% and 85% correct discrimination thresholds, which defined the difficult and easy trials respectively for the training blocks (Liu et al., 2012).

The training task used the same stimuli as the pre- and post conditions, however stimuli were presented at one of two levels, 65% and 85% accuracy, as determined at the levels described above. A QUEST procedure was used to maintain the stimuli at the appropriate performance level throughout the block (Watson & Pelli, 1983). Trials from the easy and difficult staircases were interleaved. At the start of each new block, thresholds returned to the original level determined for that day during the pre-training block. Training consisted of 12 blocks with 40 trials for each level, with a total of 80 trials within each block (960 trials per day). For half of the observers in each of the local and global conditions, trial-by-trial feedback was provided during the training blocks, as described above.

Experiment 2 - Local and Global Motion

For the motion condition, stimuli were presented for 750ms and moved a distance of 3.2' in each frame. In the pre- and post-training blocks, for the local task, all the dots moved upwards in the same direction at one of 9 angles either oriented on the vertical (0°), or tilted

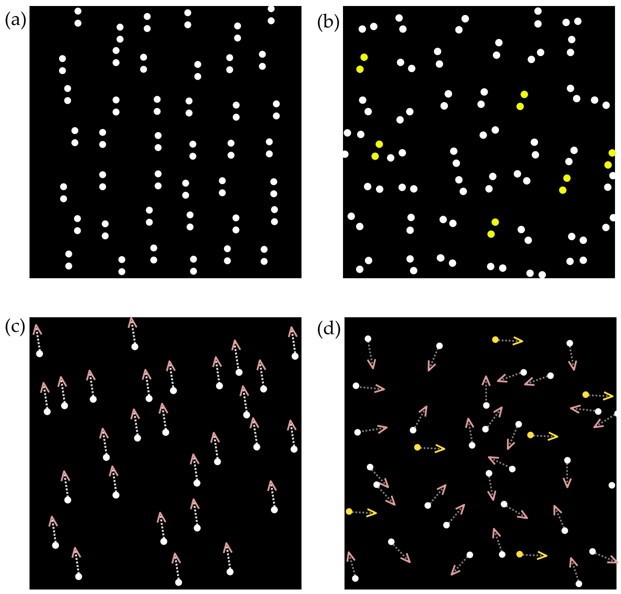

Figure 3: (a): Local Form stimuli in which all dots are oriented leftwards at 1.25° from the vertical. (b): Global form stimuli in which signal dot pairs are indicated in yellow (oriented 20° towards the right), and the remaining noise dot pairs are randomly oriented. Stimuli for the motion tasks where arrows indicate the trajectory of motion. (c) Local motion stimuli where all dots are moving upwards and are oriented leftwards at 5° from the vertical. (d) Global motion stimuli where signal dots move rightwards and noise dots move in random directions. Stimuli have been simplified (by reducing the number and increasing the size of dots) for illustration purposes.

at one of the following angles: ± 1.25°, 3.5°, 7.5°, 12.5°. Observers were required to indicate whether the dots were moving left or right from the vertical (see Figure 3 (c)). For the global task, dots were randomly designated as signal or noise. Signal dots moved to the left or right of vertical at an angle of ± 20°. Noise dots moved in a random direction uniformly selected from 0-360°. Each dot maintained the same trajectory, but had a limited lifetime of 4 frames. The next trial only began after a response was made, preceded by a fixation cross for 500 ms, and a further 500ms pause. The task was to identify the direction (either left or right of vertical) of coherent motion provided by the signal dots (see Figure 3 (d).

As in the form task, test stimuli were presented before and after training, using MOCS and responses were made using the left or right arrow on the keyboard, also containing a randomised order of trials with 30 repetitions (270 trials). A cumulative Gaussian curve was fit to responses from the daily pre-assessment each day to determine a 65% and 85% correct discrimination threshold for use in the training blocks. The training task was conducted using the same QUEST methods outlined in the form task.

Experiment 3 - Global Motion using MOCS training

For Experiment 3, pre- and post-assessments were identical to those used for experiment 2. However, to evaluate any training differences between methods of presentation (staircase or MOCS) these groups where trained using predefined randomly interleaved motion coherence levels. Training levels were identified by fitting a psychometric curve on the post-assessment results from the observers in the global motion condition from Experiment 2 to establish the mean number of dots required to achieve a percentage of correct responses ranging from 50% to 95% correct in increments of 5%. Coherence was defined at 8 levels (27, 28, 49, 78, 106, 135, 163, 210) which designated the number of signal dots present for that trial. Trials were presented randomly using MOCS for 10 repetitions at each signal level. Training consisted of 12 blocks with a total of 80 trials within each block, and 960 training trials daily.

Results

Observers completed a pre- and post-training session on each of the three days of training, which have been modelled as sessions 1-6. The daily pre- and post-assessments for the local tasks were evaluated in terms of the number of right-click responses (left vs rightward oriented) and the change over the six sessions. For the global tasks we evaluated the proportion of correct responses and the change over the six sessions. Test data were modelled using a generalised mixed effects model (detailed below) and a threshold estimate at 75% accuracy was calculated for comparison. When comparing session 1 performance across feedback groups, it is important to remember that the test-phase was identical for all groups. Thus any performance differences between feedback groups reflect random variation amongst observers.

Training was delivered using a staircase method where stimulus intensity was set to obtain an accuracy rate of 65% (difficult) and 85% (easy). In order to compare the training

Table 1: Critical $t$values for the JZS Bayes Factor Values (Rouder et al., 2009, p.232)

Supports the	Null		Alternative	
N	**10**	**3**	**1/3**	**1/10**
5		0.40	3.15	4.97
10		0.89	2.73	3.60
20		1.20	2.64	3.26
50		1.51	2.68	3.17
100	0.69	1.72	2.76	3.20
200	1.08	1.90	2.86	3.27
500	1.44	2.12	2.99	3.38

and test data, we calculated the 75% threshold for the training results (midpoint between 65% and 85%).

Finally, to determine the strength of evidence for our results, we computed a Scaled JZS Bayes Factor using a one sample Bayesian ttest with the default standardised effect size δ for the Bayes factor t-test (Cauchy: 0, 0.707). The BF_{10} was calculated for each interaction (see Table 1 for the critical values for t) using the calculator for ttests located at hich uses version 0.9.8 of the BayesFactor package (R version 3.3.2 (2016-10-31)) on i386-redhat-linux-gnu) (Rouder et al., 2009).

Experiment 1 - Local and Global Form

Local Form

For local orientation training data the daily performance average was calculated for each feedback group, showing the average degree of tilt (right or left) from the vertical required to achieve an accuracy of 65% and 85%, during training.

For the local orientation pre- and post-assessment we modelled the proportion of rightward responses as a function of stimulus orientation using a general linear mixed effects model (GLMM) using MATLAB (maximum likelihood model) with a probit link function. Groups were modelled independently. Model comparisons were undertaken for a *restricted model* (containing only the fixed effects of session and stimulus intensity) and a *full model* which also included random intercepts, and slopes for session, defined across observers. A likelihood ratio test (LRT) indicated that in both cases the full model was a significantly better fit than the restricted model (feedback group $\chi^2(3) = 121.6, p < 0.001$, and the no-feedback group $\chi^2(3) = 74.8, p < 0.001$). The model comparison is reported in Table 2 (a & b).

There was a significant improvement in performance as a result of training for the feedback group $\beta = 0.01988; t(752) = 5.9717; p < 0.001$, and the Bayes Factor value (BF) showed strong support for the alternative hypothesis, that there was an increase in the slope of the function, (indicative of learning). The analysis for the no-feedback group also showed significant improvement in performance $\beta = 0.015686; t(752) = 3.852; p < 0.001$, with BF strongly in favour of the alternative hypothesis (an increase in the slope of the

Table 2: Model comparison for local orientation discrimination where $N = 14$ for each condition (feedback and no-feedback)

(a) Feedback group *(df=752)*		Restricted Model				Full Model				Bayes Factor	
Fixed Effects	*β*	*Lower CI*	*Upper CI*	*pVal*	*β*	*Lower CI*	*Upper CI*	*pVal*	*tStat*	*BF*	*model*
Intercept	0.0740	0.0185	0.1295	0.009	0.0830	-0.0527	0.2187	0.230			
Session	-0.0021	-0.0167	0.0125	0.780	-0.0025	-0.0224	0.0174	0.804			
Stimulus	0.1509	0.1249	0.1769	$< .001^{\dagger}$	0.1567	0.1335	0.1799	$< .001^{\dagger}$			
Session × Stimulus	0.0208	0.0134	0.0281	$< .001^{*}$	0.0199	0.0133	0.0264	$< .001^{*}$	3.85	61.4	Alt
Random Effects:	Intercept	NA			0.2408						
	Session	NA			0.0287						
	Error	0.1212			0.1063						
	AIC	643.43			527.88						

(b) No-Feedback group *(df=752)*		Restricted Model				Full Model				Bayes Factor	
Fixed Effects	*β*	*Lower CI*	*Upper CI*	*pVal*	*β*	*Lower CI*	*Upper CI*	*pVal*	*tStat*	*BF*	*model*
Intercept	0.0168	-0.0371	0.0707	0.541	0.0184	-0.0741	0.1109	0.696			
Session	-0.0056	-0.0197	0.0084	0.433	-0.0066	-0.0278	0.0146	0.543			
Stimulus	0.2507	0.2190	0.2824	$< .001^{\dagger}$	0.2512	0.2222	0.2801	$< .001^{\dagger}$			
Session × Stimulus	0.0153	0.0065	0.0240	0.001	0.0157	0.0077	0.0237	$< .001^{*}$	5.51	79.1	Alt
Random Effects:	Intercept	NA			0.1498						
	Session	NA			0.0321						
	Error	0.1095			0.0998						
	AIC	720.76			651.92						

(c) Three way GLMM *(df=1504)*							Bayes Factor	
Fixed Effects	*β*	*SE*	*tStat*	*pVal*	*Lower CI*	*Upper CI*	*BF*	*model*
Intercept - Group 0	-0.0326	0.1301	-0.2510	0.802	-0.2877	0.222		
Intercept change - group 1	-0.0252	0.045	-0.5597	0.5758	-0.1134	0.0631		
Session	0.0588	0.0842	0.6985	0.485	-0.1064	0.2241		
Stimulus	0.3534	0.0386	9.1554	$< .001^{*}$	0.2777	0.4291		
Group × Session	0.0099	0.0282	0.3522	0.7248	-0.0454	0.0653	32.3	Null
Group × Stimulus	0.0157	0.0190	0.8278	0.4080	-0.0215	0.0529	24.4	Null
Session × Stimulus	-0.1042	0.0228	-4.6165	$< .001^{*}$	-0.149	-0.0599	1121.1	Alt
Group × Session × Stimulus	0.0125	0.0112	1.1167	0.2643	-0.009	0.0343	18.8	Null

Table notes : * indicates significant p-value and † performance increases significantly with increasing stimulus intensity. CI shows the 95% confidence intervals and AIC is the Akaike Information Criterion that was used in the model comparison. Plots (a) and (b) - Model comparison for the GLMM where, the intercept represents the deviation from expected performance on Day 1 (the proportion of right clicks is zero) session indicates the change in intercept over the six sessions. The Session-by-Stimulus interaction shows the change in the slope of the function across time. A significant positive change indicates evidence of learning. Bayes Factors (BF) along with their t-statistic are reported in the final column. Bayes Factor above 10 indicate strong support for the model either the Null or Alternative (Alt) which is listed in the *model* column. Plot (c) Fixed effects from the three-way GLMM to evaluate group differences. The no-feedback is coded as 0 (reference group) and the feedback groups as 1. A thresholds reflects the point of intersection between the slope and stimulus intensity, where Group-by-Stimulus is the interaction between the feedback and no-feedback groups on Day 1. Session-by-Stimulus is the change in slope across time collapsed across groups. Finally, the three-way interaction Group-by-Session-by-Stimulus is the between group interaction in the change in slope across time. BF are reported for each interaction

function). Finally a three-way analysis, with additional fixed effects of group, group-by-stimulus, group-by-session and group-by-stimulus-by-session was used to determine if the rate of learning differed between the feedback and no-feedback groups. The fixed effects are reported in full in Table 2 (c). There were no significant differences between the two groups, with the BF favouring the null for all between groups comparisons (no significant difference between groups).

Global Form (Orientation)

For the global orientation training data a daily performance average was calculated for each feedback group, showing the average number of signal dot-pairs (out of 225) required to correctly identify the direction of orientation (left or right) at performance accuracy of 65% and 85% correct.

For the global orientation pre- and post-assessment, we modelled the proportion of correct responses as a function of stimulus coherence (number of signal dot pairs), using a

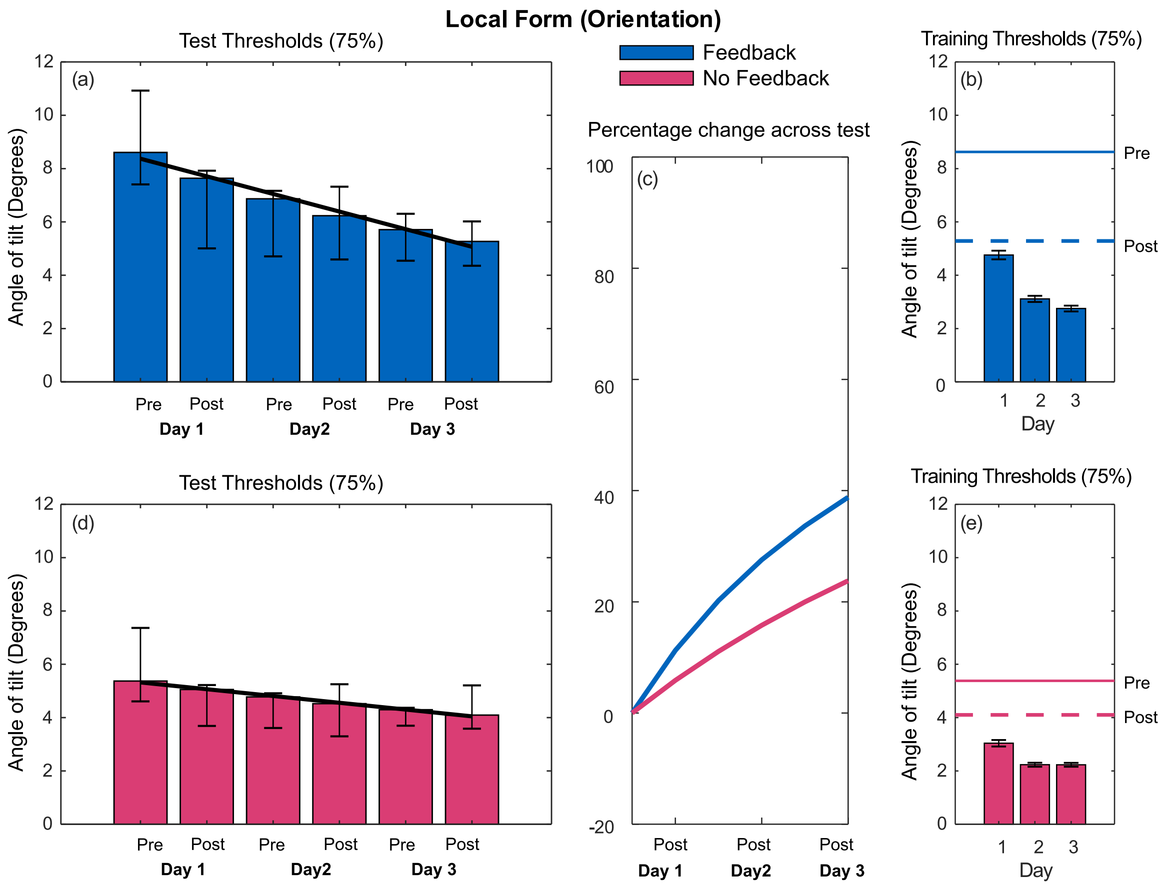

Figure 4: Estimated 75% test and training thresholds for local form (a & b) feedback group and (d & e) no-feedback group. Threshold is the degree of tilt required (from the vertical at 0°) in order to make a correct left vs right orientation decision. Error for plots (a & d) are 95% bootstrap confidence intervals. Error for (b & e) is ± standard error of the mean. Plot (c) shows the percentage of change across the 6 testing sessions.

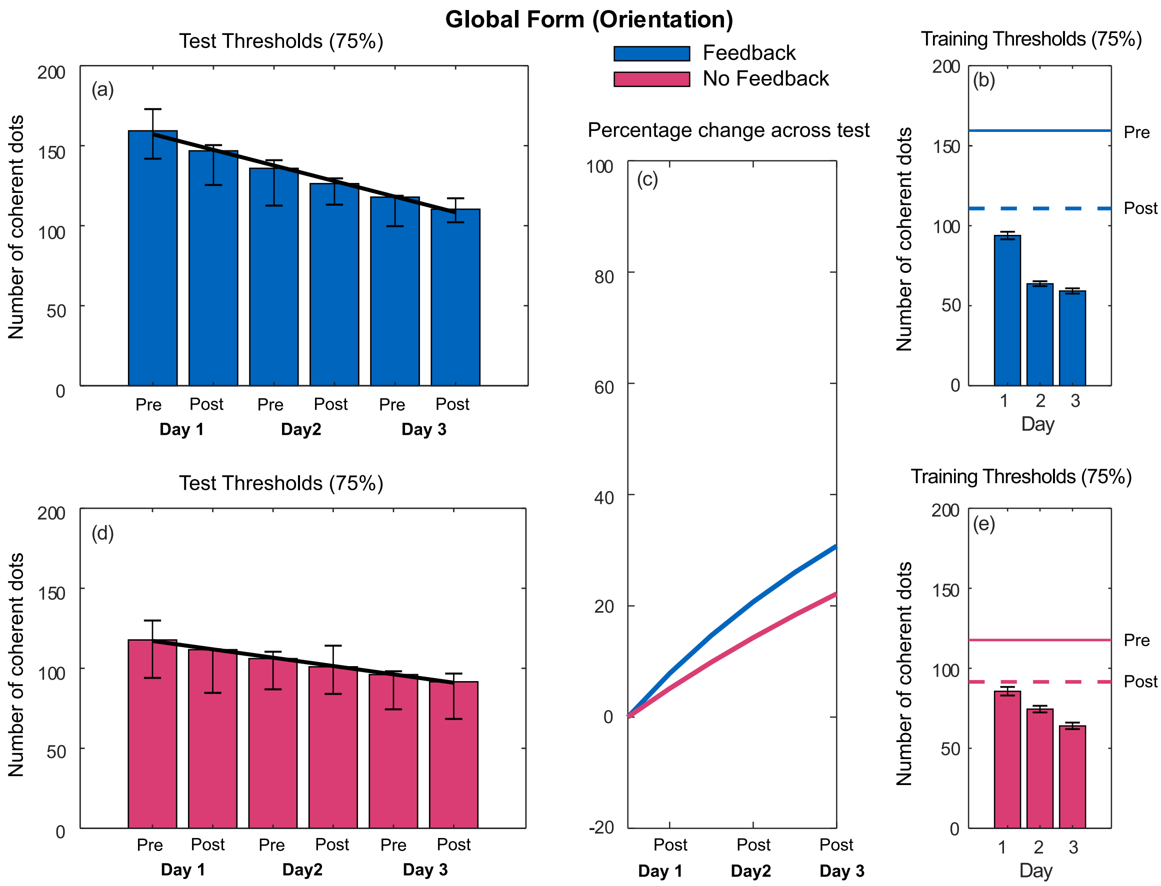

Figure 5: Estimated 75% test and training thresholds for global form (a & b) feedback group and (d & e) no-feedback group. Threshold number of coherent dot pairs (out of 225) required to make a correct left vs right orientation decision. Error for plots (a & d) are 95% bootstrap confidence intervals. Error for (b & e) is ± standard error of the mean. Plot (c) shows the percentage of change across the 6 testing sessions.

Table 3: Model comparison for global orientation discrimination where $N = 14$ for each condition (feedback and no-feedback)

(a) Feedback group *(df=752)*	Restricted Model				Full Model				Bayes Factor		
Fixed Effects	β	*Lower CI*	*Upper CI*	*pVal*	β	*Lower CI*	*Upper CI*	*pVal*	*tStat*	*BF*	*model*
(Intercept)	-1.1282	-1.2788	-0.9775	0.000	-1.1760	-1.3595	-0.9925	$< .001$			
Session	0.0190	-0.0197	0.0577	0.335	0.0218	-0.0135	0.0571	0.226			
Stimulus	0.0132	0.0112	0.0152	0.000	0.0137	0.0118	0.0155	$< .001^{\dagger}$			
Session × Stimulus	0.0010	0.0005	0.0016	0.000	0.0010	0.0005	0.0015	$< .001^{*}$	3.63	27.6	Alt
Random Effects:	Intercept	NA			0.2289						
	Error	0.1545			0.1395						
	AIC	1271.50			1213.90						

(b) No-Feedback group *(df=752)*	Restricted Model				Full Model				Bayes Factor		
Fixed Effects	β	*Lower CI*	*Upper CI*	*pVal*	β	*Lower CI*	*Upper CI*	*pVal*	*tStat*	*BF*	*model*
(Intercept)	-0.9956	-1.1687	-0.8225	0.000	-1.1672	-1.4364	-0.8980	$< .001$			
Session	0.0028	-0.0420	0.0476	0.903	0.0121	-0.0251	0.0494	0.522			
Stimulus	0.0123	0.0100	0.0145	0.000	0.0141	0.0121	0.0160	$< .001^{\dagger}$			
Session × Stimulus	0.0008	0.0002	0.0014	0.013	0.0007	0.0002	0.0012	0.009^{*}	2.59	1.14	Alt
Random Effects:	Intercept	NA			0.4286						
	Error	0.1890			0.1501						
	AIC	1455.70			1380.10						

(c) Three way GLMM *(df=1504)*							Bayes Factor	
Fixed Effects	β	*SE*	*tStat*	*pVal*	*Lower CI*	*Upper CI*		
Intercept group 0	-1.1602	0.1166	-9.9505	1.23E-22	-1.3889	-0.9315		
Intercept group 1	-0.0217	0.1654	-0.13099	0.8958	-0.34606	0.30273		
Session	0.0121	0.0183	0.66262	0.5077	-0.02373	0.047932		
Stimulus	0.0140	0.0009	14.828	1.59E-46	0.012147	0.015851		
Group × Session	0.0099	0.0262	0.37746	0.7059	-0.04144	0.061191	32.04	Null
Group × Stimulus	-0.0003	0.0014	-0.20889	0.8346	-0.00294	0.002373	33.6	Null
Session × Stimulus	0.0007	0.0003	2.6695	0.0077	0.000177	0.00116	1.01	Alt
Grp × Sesh × Stim	0.0004	0.0004	0.95184	0.3413	-0.00037	0.001073	21.9	Null

Table notes : * indicates significant p-value and † performance increases significantly with increasing stimulus intensity. CI shows the 95% confidence intervals and AIC is the Akaike Information Criterion that was used in the model comparison. Plots (a) and (b) - Model comparison for the GLMM where, the intercept represents the deviation from expected performance on Day 1 (where zero stimulus intensity is at chance performance). session indicates the change in intercept over the six sessions. The Session-by-Stimulus interaction shows the change in the slope of the function across time. A significant positive change indicates evidence of learning. Bayes Factors (BF) along with their t-statistic are reported in the final column. Bayes Factor above 10 indicate strong support for the model either the Null or Alternative (Alt) which is listed in the *model* column. Plot (c) Fixed effects from the three-way GLMM to evaluate group differences. The no-feedback is coded as 0 (reference group) and the feedback groups as 1. A thresholds reflects the point of intersection between the slope and stimulus intensity, where Group-by-Stimulus is the interaction between the feedback and no-feedback groups on Day 1. Session-by-Stimulus is the change in slope across time collapsed across groups. Finally, the three-way interaction Group-by-Session-by-Stimulus is the between group interaction in the change in slope across time. BF are reported for each interaction general linear mixed effects model using MATLAB (maximum likelihood model) with a probit link function. The feedback and no-feedback groups were modelled independently. Model comparisons were undertaken for a *restricted model* (containing only the fixed effects of session and stimulus intensity) and a *full model* which also included random intercepts and slopes defined across observers. A likelihood ratio test indicated that for both groups the full model was a significantly better fit than the restricted model (feedback group $\chi^2(1) = 59.6, p < 0.001$, and no-feedback group $\chi^2(1) = 77.6, p < 0.001$). The model comparison is reported in Table 3 (a & b). There was a significant improvement in performance for both the feedback group ($\beta = 0.0001; t(752) = -12.582; p < 0.001$), and the no-feedback group ($\beta = 0.00006; t(752) = 2.58; p = 0.009$). While BF favoured the alternative hypothesis (an increase in the slope of the function) for both groups, the evidence was strong for the group with feedback, but weak for the no feedback group. These results are plotted in figure 5 (a & d) for feedback and no-feedback groups respectively. Figure 5 (b & e) shows the training thresholds in addition to the initial pre-test threshold, and the final post-test threshold for comparison. Figure 4 (c) shows the percentage of change across the 6 testing sessions. Finally a three-way analysis with additional fixed effects of group, group-by-stimulus, group-by-session and group-by-stimulus-by-session was used to determine if the rate of learning differed between the feedback and no-feedback groups. No significant differences between the two groups were found and BF strongly supported the null hypothesis (no difference in learning between groups). BF and the full GLMM can be seen in Table 3 (c). Estimated 75% thresholds and standard errors were calculated using the same bootstrap method reported previously and are plotted in figure 5 (a & d) for feedback and no-feedback groups respectively, and see figure 8 (a & b) for the psychometric fits to the observer data. As in the local task, these results show an improvement in global orientation discrimination thresholds as a result of training, and that while both groups showed evidence of learning, evidence was stronger for the feedback group.

Experiment 2 - Local and Global Motion

Local Motion

For the local motion training data the daily performance average was calculated for each feedback group, showing the direction signal that was required to achieve an accuracy of 65% and 85% in the detection (left or rightwards) for coherent motion based on the deviation from the vertical (upwards).

While one observer in the feedback condition did not finish all three days, the data they did provide were included in the analysis. The model comparison indicated that a mixed effects model was not required, and the best fit contained only the fixed effects of session and stimulus intensity (see Table 4 (a & b)). The LRT indicated that in both cases the full model was not a better fit than the restricted model (feedback group $\chi^2(1) < 0.001, p = 0.999$, and the no-feedback group $\chi^2(1) = 2.5, p = 0.110$).

There was a significant improvement in performance (the proportion of right-key responses as a function of stimulus direction) for the no-feedback group ($\beta = 0.01344; t(752) = 3.492; p < 0.001$) and the BF strongly favoured the alternative hypothesis (a difference in the rate of learning across the sessions). However, for the feedback group there was no evidence of a change in the slope across sessions ($\beta = -0.00173; t(698) = -0.463; p = 0.644$)), and BF strongly favoured the null hypothesis (see Table 4 (a & b)). Finally there was a significant three-way interaction (see Table 4 (c)) between groups ($\beta = -0.01497; t(1450) = -2.810; p = 0.005$) indicating that learning (an increase in the slope of the function) differed significantly between groups. The BF showed anecdotal support for H_1 (a difference in learning between feedback groups). Estimated 75% thresholds and confidence intervals were calculated using the same bootstrap method reported previously and are plotted in figure 6 (a & d) for feedback and no-feedback groups respectively and see figure 6 (d & e) for the psychometric fits to the observer data.

When interpreting the difference in learning rate between the two groups, the first point to note is that the experience for all observers in the first test session was identical (see Day 1 (pre) in figure 6 (a & d)). Therefore, the difference in performance between the groups represents natural variation in performance on the task. However, in order to investigate the possibility that the difference in the initial threshold influenced the learning rate, we report here the three-way interaction between group (feedback or no feedback), session, and stimulus level, reflecting the significant negative slope for the no-feedback group (showing perceptual learning) but not for the feedback group (in Table 4 (c)). Initial performance was lower for the feedback group (Group 1) compared to the no-feedback group (Group 0), (see Group $\times$ Stimulus $\beta = 0.0457; t(1450) = 2.2671; p = 0.0235$). The analysis indicates that there was a significant difference in the rate of improvement between the two groups. Figure 6 (b & e) shows the training thresholds in addition to the initial pre-test threshold, and the final post-test threshold for comparison and figure 6 (c) shows the percentage of change across the 6 testing sessions.

Global Motion

For the global motion training data a daily performance average was calculated for each feedback group, showing the average number of signal dots (out of 225) required to correctly identify the direction of motion (left or right) at performance accuracy of 65% and 85% correct.

For the global motion pre- and post-assessment we modelled the proportion of correct responses as a function of stimulus intensity (number of coherent dot pairs), using a general linear mixed effects model using Matlab (maximum likelihood model) with a probit link function, however this model was unable to converge. The most likely explanation is because performance even at highest stimulus intensity did not reach the upper asymptote, an underlying assumption of the GLMM (Swanson & Birch, 1992). While stimulus levels for all experiments were established through pilot testing the levels, this level of performance was not matched in the global motion experiment. To accommodate the response data we used a nonlinear generalised mixed effects model (NLME) with a probit linking function. Asymptotic performance was included as a free parameter to model the variability in responses. The NLME does not provide a simple measure of significance, and accompanying Bayes Factor, for assessment of the contribution of each parameter. We thus interpret this from the 95% confidence intervals. In order to calculate these we undertook a parametric

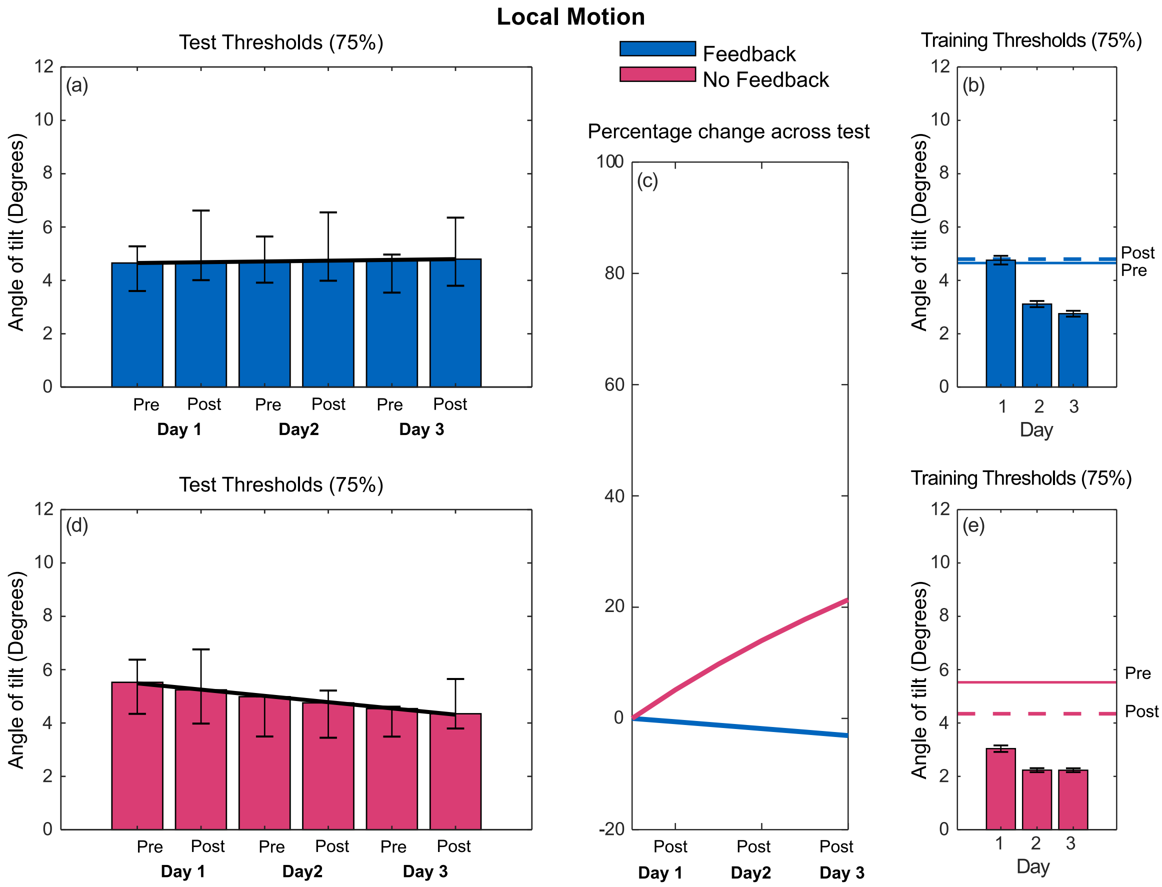

Figure 6: Estimated 75% test and training thresholds for local motion (a & b) feedback group and (d & e) no-feedback group. Threshold is the degree of tilt required (from the vertical at 0°) in order to make a correct left vs right motion direction discrimination. Error for plots (a & d) are 95% bootstrap confidence intervals. Error for (b & e) is ± standard error of the mean. Plot (c) shows the percentage of change across the 6 testing sessions.

Table 4: Model comparison for local motion discrimination where $N = 13$ in the Feedback group and $N = 14$ in the No-Feedback group

(a) Feedback group *(df=698)*	Restricted Model				Full Model				Bayes Factor		
Fixed Effects	β	*Lower CI*	*Upper CI*	*pVal*	β	*Lower CI*	*Upper CI*	*pVal*	*tStat*	*BF*	*model*
(Intercept)	-0.0247	-0.0800	0.0307	0.382	-0.0247	-0.0800	0.0307	0.382			
Session	0.0067	-0.0075	0.0209	0.354	0.0067	-0.0075	0.0209	0.354			
Stimulus	0.2899	0.2609	0.3188	$< .001^{\dagger}$	0.2899	0.2609	0.3188	$< .001^{\dagger}$			
Session × Stimulus	-0.0017	-0.0091	0.0056	0.644	-0.0017	-0.0091	0.0056	0.644	-0.46	21.1	Null
Random Effects	Intercept	NA				$< .001$					
	Error	0.0817				0.0817					
	AIC	2618.80				2620.80					

(b) No-Feedback group *(df=752)*	Restricted Model				Full Model				Bayes Factor		
Fixed Effects	β	*Lower CI*	*Upper CI*	*pVal*	β	*Lower CI*	*Upper CI*	*pVal*	*tStat*	*BF*	*model*
(Intercept)	0.0239	-0.0300	0.0779	0.384	0.0240	-0.0311	0.0790	0.393			
Session	-0.0056	-0.0197	0.0085	0.436	-0.0056	-0.0197	0.0085	0.437			
Stimulus	0.2442	0.2173	0.2711	$< .001^{\dagger}$	0.2441	0.2173	0.2710	$< .001^{\dagger}$			
Session × Stimulus	0.0132	0.0058	0.0207	0.001	0.0132	0.0058	0.0207	0.001	3.50	17.7	Alt
Random Effects	Intercept	NA			0.0219						
	Error	0.0838			0.0837						
	AIC	3025.80			3025.20						

(c) Three way GLMM *(df=1504)*							Bayes Factor	
Fixed Effects	β	*SE*	*tStat*	*pVal*	*Lower CI*	*Upper CI*		
Intercept - Group 0	0.0240	0.0274	0.8738	0.3824	-0.0298	0.0777		
Intercept change - Group 1	-0.0485	0.0398	-1.2187	0.2232	-0.1266	0.0296		
Session	-0.0056	0.0071	-0.7880	0.4308	-0.0195	0.0083		
Stimulus	0.2441	0.0135	18.0510	$< .001^{\dagger}$	0.2176	0.2707		
Group × Session	0.0123	0.0102	1.2077	0.2273	-0.0077	0.0323	16.3	Null
Group × Stimulus	0.0457	0.0201	2.2671	0.0235^{*}	0.0062	0.0852	2.61	Null
Session × Stimulus	0.0132	0.0037	3.5431	0.0004^{*}	0.0059	0.0206	15.1	Alt
Group × Session × Stimulus	-0.0150	0.0053	-2.8107	0.005^{*}	-0.0254	-0.0045	1.50	Alt

Table notes : * indicates significant p-value and † performance increases significantly with increasing stimulus intensity. CI shows the 95% confidence intervals and AIC is the Akaike Information Criterion that was used in the model comparison. Plots (a) and (b) - Model comparison for the GLMM where, the intercept represents the deviation from expected performance on Day 1 (the proportion of right clicks is zero) session indicates the change in intercept over the six sessions. The Session-by-Stimulus interaction shows the change in the slope of the function across time. A significant positive change indicates evidence of learning. Bayes Factors (BF) along with their t-statistic are reported in the final column. Bayes Factor above 10 indicate strong support for the model either the Null or Alternative (Alt) which is listed in the *model* column. Plot (c) Fixed effects from the three-way GLMM to evaluate group differences. The no-feedback is coded as 0 (reference group) and the feedback groups as 1. A thresholds reflects the point of intersection between the slope and stimulus intensity, where Group-by-Stimulus is the interaction between the feedback and no-feedback groups on Day 1. Session-by-Stimulus is the change in slope across time collapsed across groups. Finally, the three-way interaction Group-by-Session-by-Stimulus is the between group interaction in the change in slope across time. BF are reported for each interaction

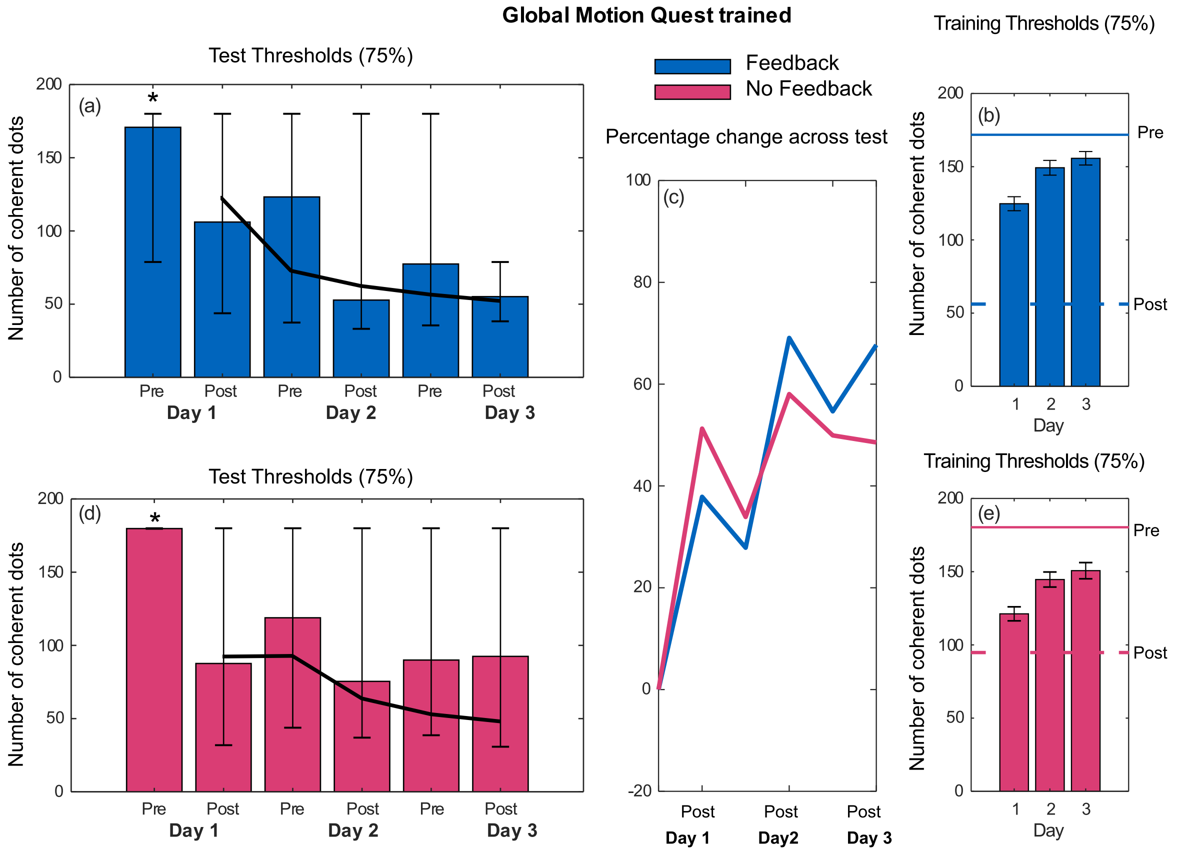

Figure 7: Estimated 75% test and training thresholds for global motion (a & b) feedback group and (d & e) no-feedback group. Threshold number of coherent dot pairs (out of 225) required to make a correct left vs right motion direction discrimination.* The model was unable to converge on a 75% threshold for day1(pre), which suggests that asymptotic performance was lower than 75%. Error for plots (a & d) are 95% bootstrap confidence intervals. Error for (b & e) is ± standard error of the mean. Plot (c) shows the percentage of change across the 6 testing sessions

bootstrap of 1000 bootstrap samples from the covariance matrix of the parameter estimates from the NLME. Estimated 75% thresholds and confidence intervals were calculated and are plotted in figure 7 (a & d) for feedback and no-feedback groups respectively. Figure 7 (b & e) shows the training thresholds in addition to the initial pre-test threshold, and the final post-test threshold for comparison and figure 7 (c) shows the percentage of change across the 6 testing sessions. Confidence intervals are illustrated graphically see figure 10 (a & b) for the psychometric fit of the NLME model, and (c & d) for 95% confidence limits for the change statistics for the asymptote and the slope respectively). Test thresholds reduced over the 3 days, and while there was no change in the slope, both groups showed a significant improvement in asymptote. There was no difference in the scale of improvement between feedback or no-feedback groups.

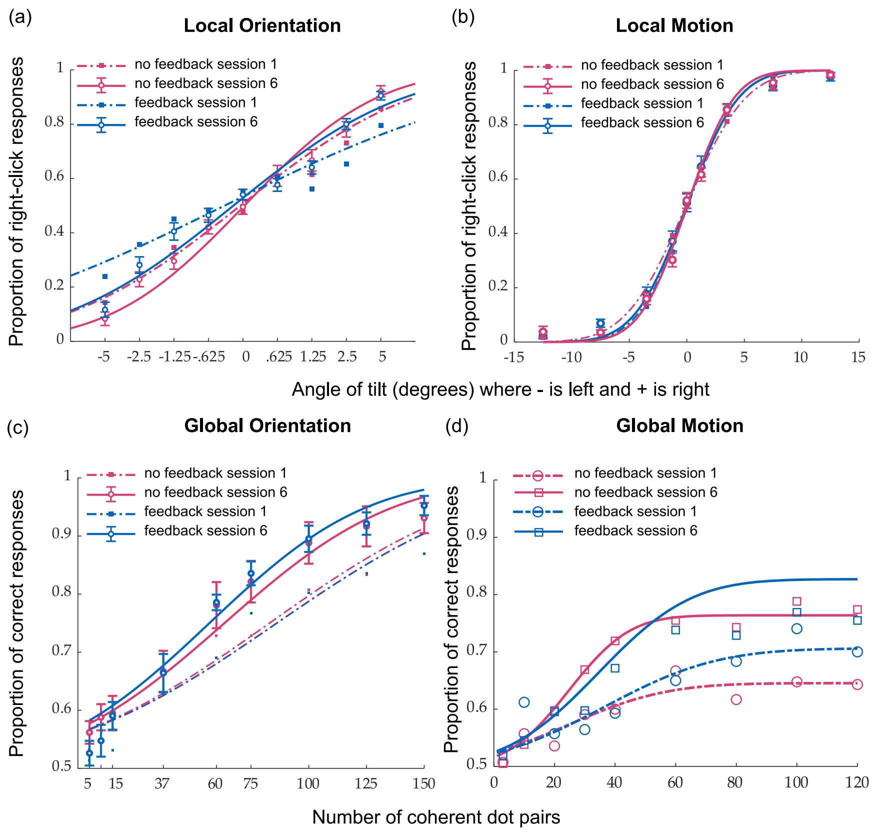

Figure 8: Psychometric fits for all groups trained using an adaptive staircase, the estimated fit has been plotted with the the means of the observer responses. For plots (a & b) the mean proportion of right-click responses are plotted as a function of of degree of tilt, where a negative stimulus depicts a tilt to the left. The global orientation and motion frames (c & d) depict the mean proportion of correct responses as a function of stimulus intensity. Where higher numbers indicate higher coherence. Groups without feedback are plotted in pink and groups without feedback are plotted in blue. While the regression contained all sessions, performance has only been plotted for the first pre-test and for the final post-test (sessions 1 and 6 respectively). Error for plots (a-c) are ± 1 SEM.

Experiment 3 - Global Motion (MOCS vs QUEST)

Global Motion (MOCS)

Unlike the previous conditions, these groups were trained using MOCS. In order to establish an estimated 75% threshold we modelled the training data using the generalised linear mixed effects model previously described, and used the bootstrap method to obtain confidence intervals of the estimation. Similar to the QUEST condition, a generalised model provided a poor fit to the response data. Thus we employed the same method as previously described, an NLME with a probit link function, and 1000 parametric bootstrap samples to estimate the 95% confidence intervals. Estimated 75% thresholds and confidence intervals were calculated and are plotted in figure 9 (a & d) for feedback and no-feedback groups respectively. Figure 9 (b & e) shows the training thresholds in addition to the initial pre-test threshold, and the final post-test threshold for comparison and figure 9 (c) shows the percentage of change across the 6 testing sessions. See figure 10 (a & b) for the psychometric fit of the NLME model, and (c & d) for 95% confidence limits for the change statistics for the asymptote and the slope respectively. Again, test thresholds reduced over the 3 days, and while there was no change in the slope, both groups showed a significant improvement in asymptote. There was no difference in the scale of improvement between feedback or no-feedback groups.

MOCS vs QUEST comparison

Only the test results were compared, and not the training data, to assess the difference of the training methods. There was a significant improvement in asymptotic performance for both groups. However, this was significantly better for the MOCS conditions, both with or without feedback.

Discussion

Understanding the links between feedback and perceptual learning is an important step in creating efficient training protocols. To understand the role of internal and external training signals in perceptual learning, we measured learning for local and global, form and motion tasks, with and without feedback, and compared the method of constant stimuli with an adaptive psychophysical technique.

We predicted that we would find robust perceptual learning with or without feedback for local conditions. In contrast, we predicted that external feedback would be more important for global conditions. This prediction was based on the increased complexity of global tasks, and findings that suggest the early (local) processing stages are more plastic than the later (global) stages (Watanabe & Sasaki, 2015). Whereas resolving local tasks is simple and relatively stimulus-driven, decisions at the global level require the segmentation and integration of stimulus elements; this added complexity may require external feedback to guide the decision making process. Finally, our predictions were also based on the assumption that learning would be more efficient when an unambiguous, external feedback signal was provided. However, since the majority of our experiments used a combination of two

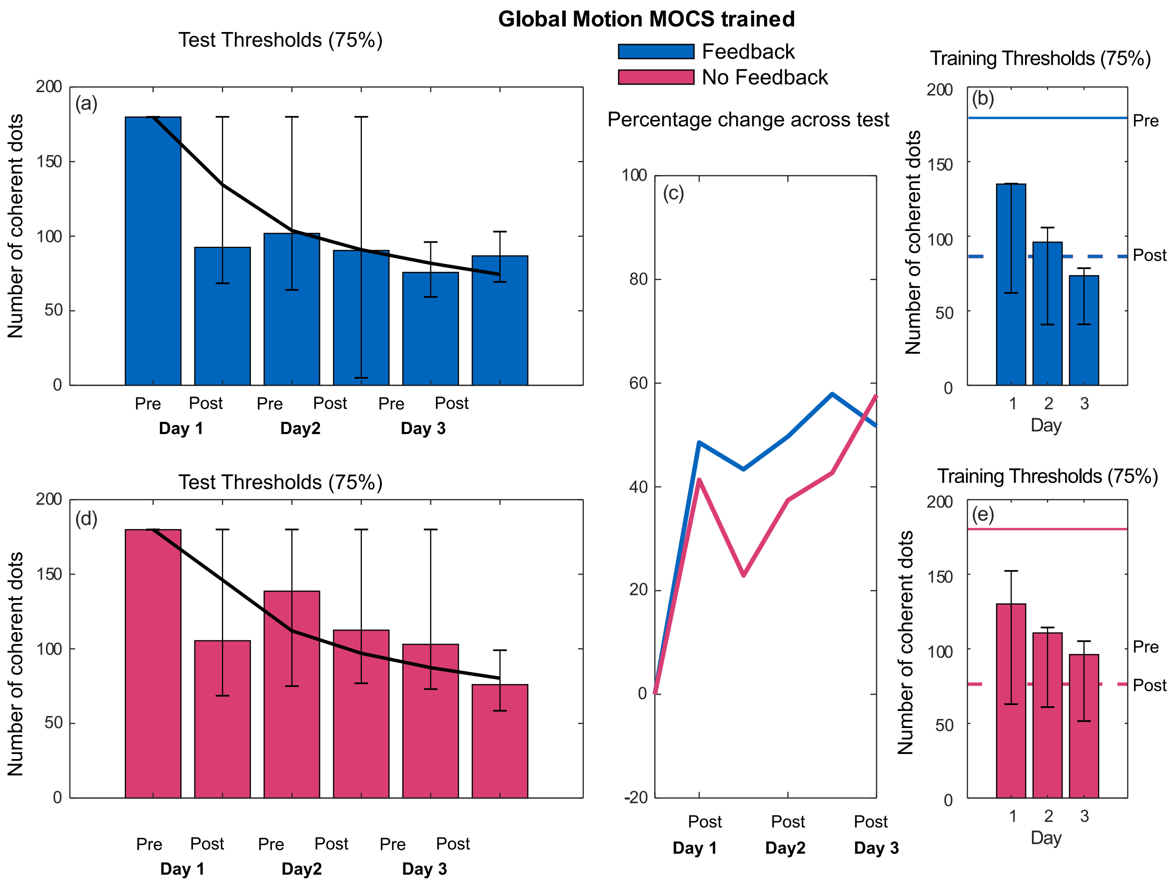

Figure 9: Estimated 75% test and training thresholds for global motion (a & b) feedback group and (d & e) no-feedback group. Threshold number of coherent dot pairs (out of 225) required to make a correct left vs right motion direction discrimination.* The model was unable to converge on a 75% threshold for day1(pre), which suggests that asymptotic performance was lower than 75%. Error for plots (a & d) are 95% bootstrap confidence intervals. Error for (b & e) is ± standard error of the mean. Plot (c) shows the percentage of change across the 6 testing sessions

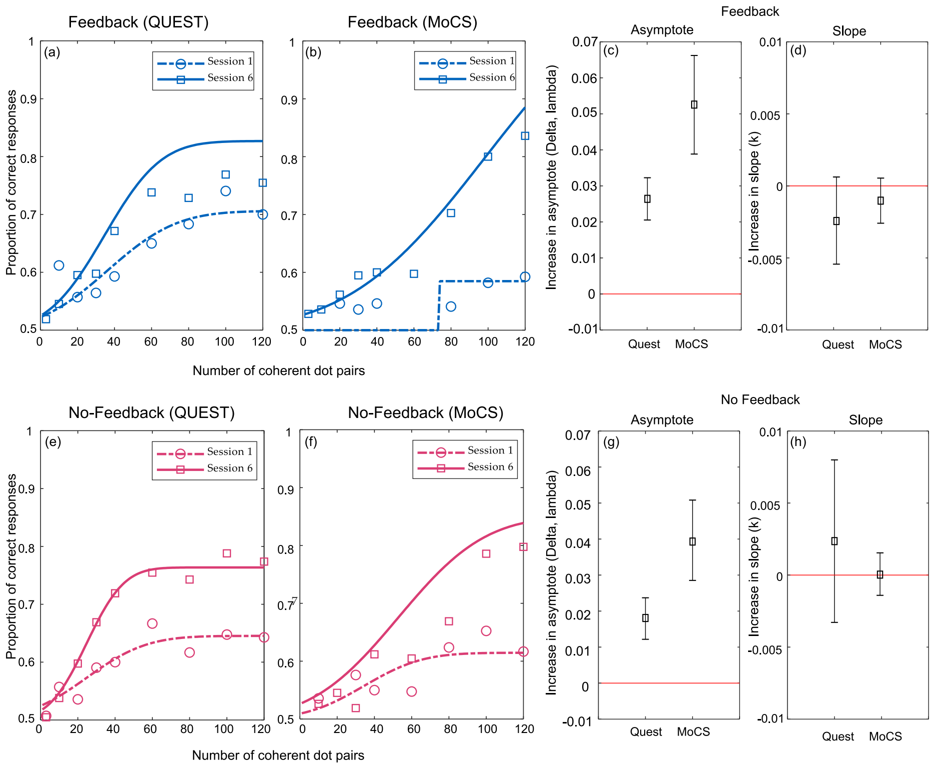

Figure 10: Comparison plots: Psychometric fit and bootstrapped confidence intervals for the global motion conditions. Groups have been merged to show performance based on the presence or absence of feedback. 95% confidence intervals that cross (include) zero are not significantly different. Those that are exclusively within a negative or positive section indicate a significant change in that direction.

stimulus levels (easy and difficult) during training, we did expect some degree of perceptual learning in all cases.

For our *local and global form* tasks, which required observers to judge the orientation of the stimulus, both groups (feedback and no-feedback) showed a statistically significant improvement across the testing sessions. The improvement across all stimulus intensities was uniform, and there were no significant between groups differences. These results suggest that observers are equally able to rely on internal or external training signals for these two tasks. In both cases, the feedback group showed a larger improvement (compared to the no-feedback group) during both the training and the test blocks. However, the final deviation (the difference between the thresholds on the final day for training and test) was better for the no-feedback group.

For *local motion,* there was a significant improvement in performance for the no-feedback group, but no change for the feedback group. While we did predict a difference between the groups, we expected that learning would occur with or without feedback, and we expected an improvement rather than reduction in the degree of learning as a result of reliable external feedback. The initial threshold for the no-feedback group was higher than the feedback group, although not significantly so. One possible interpretation of this lack of learning for the feedback group is that the higher initial level of performance did not leave scope for further improvement. In this situation the null result may reflect a ceiling effect in performance.

Global motion used the same adaptive QUEST procedure for training that was employed in all the other conditions, although thresholds increased progressively each day. There was a significant improvement across the testing sessions, most notably in asymptotic performance (indicative of improved detection at higher stimulus intensities), there were no statistically significant differences between the feedback and no-feedback groups. It is interesting that training thresholds increased with each training day, however this is likely explained by the large amount of variation in the pre-test thresholds.

Finally, for *global motion (MOCS trained),* there was a significant improvement in asymptotic performance, and no statistically significant differences between the feedback and no-feedback groups.

MOCS vs Adaptive staircase

The asymptotic improvement was significantly better for MOCS trained groups than for the QUEST trained groups for both feedback conditions. Again, this pattern of results was not what we predicted. Based on the robust learning found when only two stimulus levels were presented during training (Liu et al., 2012), and the lack of learning found with the multiple stimulus levels used in the method of constant stimuli (Seitz et al., 2006; Asher et al., 2019), we predicted weaker learning for the group trained using MOCS and without external feedback.

We had speculated that MOCS may provide a less reliable training signal than when only two levels of difficulty are provided, due to the increased complexity of metacognitive judgements of perceptual confidence in the former case (Zylberberg et al., 2014). In a previous study, we found perceptual learning for a global motion task with MOCS training only when external feedback was provided (experiment 1 (Asher et al., 2019)). Secondly when feedback was removed performance returned to baseline (experiment 2 (Asher et al., 2019)) in some conditions. Thus, in this group we anticipated that with neither a reliable external nor internal training signal would be available. However, the lack of the predicted difference between the training protocols when compared directly in the current study, does not support this position.

The training thresholds for both global motion feedback groups (quest training) increased across the three days. The increasing threshold (reported as 75% accuracy) suggests that, during training, QUEST trained observers required *more* signal dots each day to maintain an average performance accuracy of 85% and 65% (for easy and difficult trials respectively), compared to the previous day(s). This may suggest that the condition was particularly difficult, and only having two training levels (65% and 85% accuracy) created less certainty. QUEST training levels were based on the training accuracy levels used in Liu et al. (2012), however there is scope to further understand the quantifier for 'easy' or 'difficult'. Importantly, these levels were sufficient for eliciting learning in the test phase, It should also be noted that neither Liu et al. (2012) nor Seitz et al. (2006) included baseline tests before and after training, but based their conclusions on changes in performance during the training itself. Our pre- and post- training blocks were included to provide a more robust test of whether any improvements in performance were retained robustly outside of the training period. In contrast, the MOCS trained groups' performance was similar throughout the experiment, with both training and test thresholds reducing over the three days. However, the MOCS trained groups were exposed to a broad and consistent range of difficulties in training. The MOCS trained groups were trained on persistent levels stimuli with a full range of levels, including extremely high coherence levels (up to 210 of 225 dots or 93% coherence).

Overall, there may have been a boost to learning within the training blocks for QUEST groups, but this was not maintained when changing the paradigm to MOCS for testing. Interestingly in the global motion conditions, training using a full range of randomly interleaved stimulus levels covering a range of difficulties facilitated learning more than when trained with QUEST.

Summary

In general, training using an adaptive staircase at predefined levels of accuracy, interleaving easy and difficult trials, achieved lower thresholds, with and without feedback. While our initial hypothesis had predicted that performance and improvement would be better with trial-by-trial feedback our results suggest that this is not always the case. Generally, these results suggest that when observers are provided with implicit and explicit feedback, they are more reliant on explicit feedback. There is still the benefit from implicit feedback, albeit not consistent across all conditions. These findings are consistent with studies that propose a disassociation between perceptual learning as an increase in perceptual sensitivity or a change to the decision criterion (Aberg & Herzog, 2012; Jones et al., 2015) or transient learning such as priming (Lin et al., 2017; Lin & Murray, 2014). The change to the criterion is thought to depend on explicitly provided cues, such as trial-by-trial feedback. Furthermore, there is evidence that the decision criterion can change, or be purposely manipulated, from trial to trial (Jones et al., 2015; Aberg & Herzog, 2012; Herzog et al., 2006; Shibata et al., 2009). Reverse feedback, usually for a subthreshold stimulus-level, has been used to change an observer's subjective criterion decision (Herzog et al., 2006; Liu et al., 2015) and create transient biases. This would result in short-term learning that is not maintained over time, and may explain results that suggest removing feedback, after training with feedback, "freezes" performance. Research has shown that improvement slows or remains consistent (Herzog & Fahle, 1997), or even return performance to baseline (prior to any training) (Asher et al., 2019). This suggests that learning with feedback comes with a "feedback penalty", when feedback is removed. Where decision learning outweighs sensitivity learning, improvement is transient. However, by interleaving easy and difficult trials, both types of learning appeared to co-exist, albeit without any explicit benefit.

This experiment has not resolved the question raised by the absence of learning in some studies (Asher et al., 2019; Seitz et al., 2006). In the current study, there was no deficit from training on interleaved stimulus intensities, when using MOCS. Furthermore, overall performance was better than the QUEST trained groups, both with and without feedback. One difference across the experiments was that while Asher et al. (2019) used used an equivalent-noise task, this study used a discrete-noise paradigm. A discrete-noise paradigm was chosen in order for all our tasks to be as similar across conditions as possible. In a discrete-noise paradigm, signal direction is usually defined beforehand. The observer will know to attend to a set of directions (up/down or left/right etc). There will be a clear and expected signal at the local and global levels. Furthermore, since there is an expectation for the signal it is possible that learning may occur more readily as a result of up-weighting for those receptive fields tuned to those directions, based on (i) probabilistic inference at the sensory processing stages (Bejjanki et al., 2011) and (ii) the reentrant (feedback) connections from the global to local levels (Romei et al., 2016). In contrast, performance on the equivalent-noise tasks requires summing the dots across all the directions and there would be spatial overlap in neural populations responding at the local levels. Thus, neither probabilistic inference nor the up-weighting of neural populations would aide in efficiently encoding the mean direction creating a level of uncertainty in prediction and inference. This may indicate that for learning, equivalent-noise paradigms require feedback for learning, but that removing feedback may reduce performance as found in Asher et al. (2019).

We acknowledge that in some conditions there was variation in the starting thresholds which may have influenced the capacity for learning. However, we chose a repeated measures design to measure learning, which determines the rate in performance across the learning period and the rate of change, regardless of initial level of performance. The nature of our protocol, in which testing and training were integrated with experiments, allowed us to capture the immediate as well as maintained effects of learning. However, this also means that it would not have been possible to pre-allocate observers to groups (so as to remove mean differences) without additional testing that would, in itself, have provided an additional opportunity for perceptual learning.

Finally, there were performance differences between the global motion and global form conditions (direction across space) that may be explained by the additional feature in motion (direction across space and time).

Conclusions

Our findings suggest that improvement for form tasks occurred in all conditions and was robust and similar with or without feedback. For motion tasks,for our stimulus levels, performance was generally high for local motion and but less so for global motion. Ultimately, internal reinforcement is as effective as external feedback when difficult trials are interleaved with easy exemplars and use a discrete-noise paradigm. Despite the theoretical considerations that drove the design of the experiments reported here, there do not appear to be distinct situations in which external feedback is needed for perceptual learning when easy and difficult trials are interleaved. That is, learning was observed for local and global tasks; for form and motion; whether feedback was present or absent; and when using an adaptive training procedure with two levels of difficulty, or when using MOCS. Variations in the importance of feedback both within the current study and across the broader literature appear not to follow any obvious pattern, and may instead simply reflect variability across observers and studies, rather than providing any important theoretical insights.

References

Aberg, K. C., & Herzog, M. H. (2009). Interleaving bisection stimuli - randomly or in sequence - does not disrupt perceptual learning, it just makes it more difficult. *Vision Research*, *49*, 2591–2598. doi:10.1016/j.visres.2009.07.006.

Aberg, K. C., & Herzog, M. H. (2012). Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *Journal of Vision*, *12*, 10–10. doi:10.1167/12.3.3.

Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in Cognitive Sciences*, *11*, 458–465. doi:10.1016/j.tics.2007.08.015.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Letters to Nature*, *387*, 401–406. doi:10.1038/387401a0.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*, 457–464. doi:10.1016/j.tics.2004.08.011.

Amano, K., Edwards, M., Badcock, D. R., & Nishida, S. (2009). Spatial-frequency tuning in the pooling of one- and two-dimensional motion signals. *Vision Research*, *49*, 2862–2869. doi:10.1016/j.visres.2009.08.026.

Asher, J., Romei, V., & Hibbard, P. (2019). Spatial Frequency Tuning and Transfer of Perceptual Learning for Motion Coherence Reflects the Tuning Properties of Global Motion Processing. *Vision*, *3*, 44. doi:10.3390/vision3030044.

Badcock, D. R., Almeida, R. A., & Dickinson, J. E. (2013). Detecting global form: separate processes required for Glass and radial frequency patterns. *Frontiers in computational neuroscience*, *7*, 53. doi:10.3389/fncom.2013.00053.

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science*, *218*, 697–698. doi:10.1126/science.7134968.

Bejjanki, V. R., Beck, J. M., Lu, Z.-L., & Pouget, A. (2011). Perceptual learning as improved probabilistic inference in early sensory areas. *Nature Neuroscience*, *14*, 642–50. URL: http://linkinghub.elsevier.com/retrieve/pii/S0733861903000963. doi:10.1016/S0733-8619(03)00096-3.

Bex, P. J., & Dakin, S. C. (2002). Comparison of the spatial-frequency selectivity of local and global motion detectors. *Journal of the Optical Society of America*, *19*, 670–7. doi:10.1364/JOSAA.19.000670.

Braddick, O. J. (1993). Segmentation versus integration in visual motion processing. *Trends in Neurosciences*, *16*, 263–268. doi:10.1016/0166-2236(93)90179-P.

Braddick, O. J., Brien, J. M. D. O., Atkinson, J., Turner, R., June, C. B., & S, F. (2000). Form and motion coherence activate independent , but not dorsal / ventral segregated networks in the human brain. *Human Brain Mapping*, *09*, 3–5.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433–436.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1993). Responses of neurons in macaque MT to stochiastic motion signals.

Burr, D., & Thompson, P. (2011). Motion psychophysics: 1985-2010. *Vision Research*, *51*, 1431–1456. doi:10.1016/j.visres.2011.02.008.

Callaway, E. M. (2004). Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Networks*, *17*, 625–632. doi:10.1016/j.neunet.2004.04.004.

Chu, M. W., Li, W. L., & Komiyama, T. (2016). Balancing the Robustness and Efficiency of Odor Representations during Learning. *Neuron*, *92*, 174–186. URL: http://dx.doi.org/10.1016/j.neuron.2016.09.004. doi:10.1016/j.neuron.2016.09.004.

Chung, C. Y., & Khuu, S. K. (2014). The processing of coherent global form and motion patterns without visual awareness. *Frontiers in Psychology*, *5*, 1–11. doi:10.3389/fpsyg.2014.00195.

Cumming, B. G., Shapiro, S. E., & Parker, A. J. (1998). Disparity detection in anticorrelated stereograms. *Perception*, *27*, 1367–1377. doi:10.1068/p271367.

Dakin, S. C., Mareschal, I., & Bex, P. J. (2005). Local and global limitations on direction integration assessed using equivalent noise analysis. *Vision Research*, *45*, 3027–3049. doi:10.1016/j.visres.2005.07.037.

Deveau, J., Ozer, D. J., & Seitz, A. R. (2014). Improved vision and on-field performance in baseball through perceptual learning. *Current Biology*, *24*, R146–R147. doi:10.1016/j.cub.2014.01.004.

Dinse, H. R., Ragert, P., Pleger, B., Schwenkreis, P., & Tegenthoff, M. (2003). Pharmacological Modulation of Perceptual Learning and Associated Cortical Reorganization TL - 301. *Science*, *301 VN -*, 91–94. URL: http://dx.doi.org/10.1126/science.1085423. doi:10.1126/science.1085423.

Dobres, J., & Watanabe, T. (2012). Response feedback triggers long-term consolidation of perceptual learning independently of performance gains. *Journal of Vision*, *12*, 9–9. doi:10.1167/12.8.9.

Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research*, *39*, 3197–3221. doi:10.1016/B978-012375731-9/50082-3.

Dosher, B. A., & Lu, Z. L. (2009). Hebbian Reweighting on Stable Representations in Perceptual Learning. *Learning & Perception*, *1*, 37–58. doi:10.1556/LP.1.2009.1.4.

Dosher, B. A., & Lu, Z.-L. (2017). Visual Perceptual Learning and Models. *Annual Review of Vision Science*, *3*, 1–24. doi:10.1146/annurev-vision-102016-061249.

Fahle, M. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology*, *15*, 154–160. doi:10.1016/j.conb.2005.03.010.

Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research*, *33*, 397–412. doi:10.1016/0042-6989(93)90094-D.

Felleman, D. J., & Van Essen, D. C. (1987). Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *Journal of neurophysiology*, *57*, 889–920. URL: http://www.ncbi.nlm.nih.gov/pubmed/3585463.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47. doi:10.1093/cercor/1.1.1.

Furlan, M., & Smith, A. T. (2016). Global Motion Processing in Human Visual Cortical Areas V2 and V3. *Journal of Neuroscience*, *36*, 7314–7324. URL: http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0025-16.2016. doi:10.1523/JNEUROSCI.0025-16.2016.

Garcia, A., Kuai, S. G., & Kourtzi, Z. (2013). Differences in the time course of learning for hard compared to easy training. *Frontiers in Psychology*, *4*, 1–8. doi:10.3389/fpsyg.2013.00110.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The Neural Basis of Perceptual Learning Review. *Neuron*, *31*, 681–697. doi:10.1016/S0896-6273(01)00424-X.

Glass, L. (1969). Moire effect from random dots. *Nature*, *223*, 578. doi:10.1038/224488a0.

Graziano, M. S., Andersen, R. A., & Snowden, R. J. (1994). Tuning of MST neurons to spiral motions. *Journal of Neuroscience*, *14*, 54–67. doi:10.1523/jneurosci.14-01-00054.1994.

Gu, X., & FitzGerald, T. H. (2014). Interoceptive inference: Homeostasis and decision-making. *Trends in Cognitive Sciences*, *18*, 269–270. URL: http://dx.doi.org/10.1016/j.tics.2014.02.001. doi:10.1016/j.tics.2014.02.001.

Guggenmos, M., Wilbertz, G., Hebart, M. N., & Sterzer, P. (2016). Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *eLife*, *5*, 1–19. doi:10.7554/eLife.13388.

Harrison, R. V., Gordon, K. A., & Mount, R. J. (2005). Is there a critical period for cochlear implantation in congenitally deaf children? Analyses of hearing and speech perception performance after implantation. *Developmental Psychobiology*, *46*, 252–261. doi:10.1002/dev.20052.

Herzog, M. H., Ewald, K. R. F., Hermens, F., & Fahle, M. (2006). Reverse feedback induces position and orientation specific changes. *Vision Research*, *46*, 3761–3770. doi:10.1016/j.visres.2006.04.024.

Herzog, M. H., & Fahle, M. (1997). The Role of Feedback in Learning a Vernier Discrimination Task. *Vision Research*, *37*, 2133–2141. doi:http://dx.doi.org/10.1016/S0042-6989(97)00043-6.

Herzog, M. H., & Fahle, M. (1998). Modeling perceptual learning: difficulties and how they can be overcome. *Biological cybernetics*, *78*, 107–117. doi:10.1007/s004220050418.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional acrhitecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of neurophysiology*, *28*, 229–289.

Huxlin, K. R., Martin, T., Kelly, K., Riley, M., Friedman, D. I., Burgin, W. S., & Hayhoe, M. (2009). Perceptual relearning of complex visual motion after V1 damage in humans. *The Journal of Neuroscience*, *29*, 3981–3991. doi:10.1523/JNEUROSCI.4882-08.2009.

Jones, P. R., Moore, D. R., & Amitay, S. (2015). The role of response bias in perceptual learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *2015*, 1456–1470. doi:10.1037/xlm0000111.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 4966–4970. doi:DOI10.1073/pnas.88.11.4966.

Kellman, P. J. (2013). Adaptive and Perceptual Learning Technologies in Medical Education and Training. *Military Medicine*, *178*, 98–106. URL: http://publications.amsus.org/doi/abs/10.7205/MILMED-D-13-00218. doi:10.7205/MILMED-D-13-00218.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., & others (2007). Whats new in Psychtoolbox-3. *Perception*, *36*, 1.

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: gaze-tracking study. *Radiology*, *242*, 396–402. doi:10.1148/radiol.2422051997.

Lamme, V. A. F. (2003). Recurrent corticocortical interactions in neural disease. *Archives of Neurology*, *60*, 178–184. doi:10.1001/archneur.60.2.178.

Lennie, P. (1998). Single units and visual cortical organization. *Perception*, *27*, 889–935. doi:10.1068/p270889.

Levi, A., Shaked, D., Tadin, D., & Huxlin, K. R. (2015). Is improved contrast sensitivity a natural consequence of visual training? *Journal of Vision*, *14*, 1158–1158. doi:10.1167/14.10.1158.

Lin, Z., Dosher, B. A., & Lu, Z.-L. (2017). Mixture of easy trials enables transient and sustained perceptual improvements through priming and perceptual learning. *Scientific Reports*, *7*, 7421. URL: http://www.nature.com/articles/s41598-017-06989-0. doi:10.1038/s41598-017-06989-0.

Lin, Z., & Murray, S. O. (2014). Priming of awareness or how not to measure visual awareness. *Journal of Vision*, *14*, 1–17. URL: http://www.journalofvision.org/lookup/doi/10.1167/14.1.27. doi:10.1167/14.1.27.

Liu, J., Dosher, B. A., & Lu, Z.-L. (2014). Modeling trial by trial and block feedback in perceptual learning. *Vision Research*, *99*, 46–56. URL: https://linkinghub.elsevier.com/retrieve/pii/S0042698914000030. doi:10.1016/j.visres.2014.01.001.

Liu, J., Dosher, B. A., & Lu, Z.-L. (2015). Augmented Hebbian reweighting accounts for accuracy and induced bias in perceptual learning with reverse feedback. *Journal of vision*, *15*, 10. doi:10.1167/15.10.10.

Liu, J., Lu, Z.-L., & Dosher, B. A. (2010). Augmented Hebbian reweighting: interactions between feedback and training accuracy in perceptual learning. *Journal of vision*, *10*, 29. doi:10.1167/10.10.29.

Liu, J., Lu, Z. L., & Dosher, B. A. (2012). Mixed training at high and low accuracy levels leads to perceptual learning without feedback. *Vision Research*, *61*, 15–24. doi:10.1016/j.visres.2011.12.002.

Liu, Z., & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision Research*, *40*, 97–109. doi:10.1016/S0042-6989(99)00140-6.

Livingstone, M., & Hubel, D. H. (1988). Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception. *Science*, *240*, 740–749.

Mamassian, P. (2016). Visual confidence. *Annual Review of Vision Science*, *2*, 459–481.

Mansouri, B., & Hess, R. F. (2006). The global processing deficit in amblyopia involves noise segregation. *Vision Research*, *46*, 4104–4117. doi:10.1016/j.visres.2006.07.017.

Maunsell, J., Nealey, T., & DePriest, D. (1990). Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. *Journal of Neuroscience*, *10*, 3323–3334. URL: http://www.jneurosci.org/cgi/content/abstract/10/10/3323. doi:10.1523/JNEUROSCI.10-10-03323.1990.

Mikami, A., Newsome, W. T., & Wurtz, R. H. (1986). Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. *Journal of neurophysiology*, *55*, 1328–39. URL: http://www.ncbi.nlm.nih.gov/pubmed/3734858.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat visual cortex. *J. Physiol.*, *283*, 101–120.

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

Nishida, S. (2011). Advancement of motion psychophysics : Review 2001 2010. *Science*, *11*, 1–53. doi:10.1167/11.5.11.Introduction.

Ostwald, D., Lam, J. M., Li, S., & Kourtzi, Z. (2008). Neural Coding of Global Form in the Human Visual Cortex. *Journal of Neurophysiology*, *99*, 2456–2469. doi:10.1152/jn.01307.2007.

Ott, T., Masset, P., & Kepecs, A. (2018). The Neurobiology of Confidence: From Beliefs to Neurons. *Cold Spring Harbor Symposia on Quantitative Biology*, *83*, 9–16. doi:10.1101/sqb.2018.83.038794.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*, 437–442.

Petrov, A. A., Dosher, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*, *112*, 715–743. doi:10.1037/0033-295X.112.4.715.

Petrov, A. A., Dosher, B. A., & Lu, Z. L. (2006). Perceptual learning without feedback in non-stationary contexts: Data and model. *Vision Research*, *46*, 3177–3197. doi:10.1016/j.visres.2006.03.022.

Prins, N., & Kingdom, F. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data.

Romei, V., Chiappini, E., Hibbard, P., & Avenanti, A. (2016). Empowering Reentrant Projections from V5 to V1 Boosts Sensitivity to Motion. *Current Biology*, *26*, 2155–2160. doi:10.1016/j.cub.2016.06.009.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, *16*, 225–237. doi:10.3758/PBR.16.2.225.

Rudolph, K., & Pasternak, T. (1999). Transient and permanent deficits in motion perception after lesions of cortical areas MT and MST in the macaque monkey. *Cerebral Cortex*, *9*, 90–100. doi:10.1093/cercor/9.1.90.

Scase, M. O., Braddick, O. J., & Raymond, J. E. (1996). What is Noise for the Motion System ? *Vision Research*, *36*, 2579–2586.

Seitz, A. R., Nanez, J. E., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision*, *6*, 9. URL: http://jov.arvojournals.org/article.aspx?doi=10.1167/6.9.9. doi:10.1167/6.9.9.

Shibata, K., Yamagishi, N., Ishii, S., & Kawato, M. (2009). Boosting perceptual learning by fake feedback.

Vision Research, *49*, 2574–2585. doi:10.1016/j.visres.2009.06.009.

Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception & psychophysics*, *52*, 582–588. doi:10.3758/BF03206720.

Sillito, A. M., Cudeiro, J., & Jones, H. E. (2006). Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends in Neurosciences*, *29*, 307–316. doi:10.1016/j.tins.2006.05.001.

Simoncelli, E. P., & Heeger, D. J. (1998). A Model of Neuronal Responses in Visual Area MT. *Vision Research*, *38*, 743–761. doi:S0042698997001831[pii].

Sterkin, A., Levy, Y., Pokroy, R., Lev, M., Levian, L., Doron, R., Yehezkel, O., Fried, M., Frenkel-Nir, Y., Gordon, B., & Polat, U. (2018). Vision improvement in pilots with presbyopia following perceptual learning. *Vision Research*, *152*, 61–73. URL: http://dx.doi.org/10.1016/j.visres.2017.09.003. doi:10.1016/j.visres.2017.09.003.

Swanson, W. H., & Birch, E. E. (1992). Extracting thresholds from noisy psychophysical data. *Perception & Psychophysics*, *51*, 409–422. doi:10.3758/BF03211637.

Talluri, B. C., Hung, S.-C., Seitz, A. R., & Seriès, P. (2015). Confidence-based integrated reweighting model of task-difficulty explains location-based specificity in perceptual learning. *Journal of vision*, *15*, 17. URL: http://jov.arvojournals.org/article.aspx?articleid=2480172&resultClick=1. doi:10.1167/15.10.17.

Tibber, M. S., Kelly, M. G., Jansari, A., Dakin, S. C., & Shepherd, A. J. (2014). An inability to exclude visual noise in migraine. *Investigative Ophthalmology and Visual Science*, *55*, 2539–2546. doi:10.1167/iovs.14-13877.

Vaina, L. M., Belliveau, J. W., Roziers, E. B. d., & Zeffiro, T. A. (1998). Neural systems underlying learning and representation of global motion. *Proceedings of the National Academy of Sciences*, *95*, 12657–12662. doi:10.1073/pnas.95.21.12657.

Vaina, L. M., Sundareswaran, V., & Harris, J. G. (1995). Learning to ignore: psychophysics and computational modeling of fast learning of direction in noisy motion stimuli. *Cognitive Brain Research*, *2*, 155–163. doi:10.1016/0926-6410(95)90004-7.

Watanabe, T., Náñez, J. E., Koyama, S., Mukai, I., Liederman, J., & Sasaki, Y. (2002). Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task. *Nature neuroscience*, *5*, 1003–1009. doi:10.1038/nn915.

Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: Toward a comprehensive theory. *Annual review of psychology*, *3*, 197–221. doi:10.1146/annurev-psych-010814-015214.Perceptual.

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

Wilson, D. A., & Stevenson, R. J. (2003). Olfactory perceptual learning: The critical role of memory in odor discrimination. *Neuroscience and Biobehavioral Reviews*, *27*, 307–328. doi:10.1016/S0149-7634(03)00050-2.

Wilson, H. R., & Wilkinson, F. (1998). Detection of global structure in Glass patterns: Implications for form vision. *Vision Research*, *38*, 2933–2947. doi:10.1016/S0042-6989(98)00109-6.

Zeki, S. (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*, 549–573. URL: http://doi.wiley.com/10.1113/jphysiol.1974.sp010452. doi:10.1113/jphysiol.1974.sp010452.

Zeki, S. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature*, *274*, 423–428. doi:10.1038/274423a0.

Zeki, S. (2003). Functional Specialisation in the Visual Brain: Probable and Improbable Visual Areas. In R. S. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, C. J. Price, S. Zeki, J. T. Ashburner, & W. D. Penny (Eds.), *Human Brain Function: Second Edition* chapter 9. (pp. 161–170). Academic Press. (2nd ed.). URL: https://www.sciencedirect.com/science/article/pii/B978012264841050010X. doi:10.1016/B978-012264841-0/50010-X.

Zylberberg, A., Roelfsema, P. R., & Sigman, M. (2014). Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness and Cognition*, *27*, 246–253. URL: http://dx.doi.org/10.1016/j.concog.2014.05.012. doi:10.1016/j.concog.2014.05.012.