

# Rapidly learned stimulus expectations alter perception of motion

**Matthew Chalk\***

School of Informatics, Edinburgh University  
Edinburgh, UK



**Aaron R. Seitz\***

Department of Psychology, University of California, Riverside  
Riverside, CA, USA



**Peggy Seriès**

School of Informatics, Edinburgh University  
Edinburgh, UK



Expectations broadly influence our experience of the world. However, the process by which they are acquired and then shape our sensory experiences is not well understood. Here, we examined whether expectations of simple stimulus features can be developed implicitly through a fast statistical learning procedure. We found that participants quickly and automatically developed expectations for the most frequently presented directions of motion, and that this altered their perception of new motion directions, inducing attractive biases in the perceived direction as well as visual hallucinations in the absence of a stimulus. Further, the biases in motion direction estimation that we observed were well explained by a model that accounted for participants' behaviour using a Bayesian strategy, combining a learned prior of the stimulus statistics (the expectation) with their sensory evidence (the actual stimulus) in a probabilistically optimal manner. Our results demonstrate that stimulus expectations are rapidly learned and can powerfully influence perception of simple visual features.

**Keywords:** expectation, motion perception, Bayesian, attention, psychophysics

\* equal contribution authors

## Introduction

As well as depending on the sensory input that we receive, our perception of the world is shaped by our expectations. These expectations can be manipulated quickly through sensory cues or experimentalists' instructions (Posner, Snyder, & Davidson, 1980; Sterzer, Frith, & Petrovic, 2008), or more slowly, based on the statistics of previous sensory inputs. For example, in complex scenes, objects are recognized faster and more accurately when they are contextually appropriate to the visual scene as a whole: when presented with an image of a kitchen, people are better at recognizing a loaf of bread than a drum (Bar, 2004). In other words, we learn from past experience which objects are expected within the context of a particular visual scene, and our perceptual sensitivity for these objects is increased accordingly.

Indeed, it has been shown extensively that expectations modulate perceptual performance. When visual cues are used to inform participants the location that a stimulus is most likely to appear, their perceptual sensitivity for stimuli presented at this location is increased. This results in decreased reaction times, decreased detection thresholds and increased sensitivity for discrimination of features such as orientation, form or brightness for stimuli presented at the

expected location (Doherty, Rao, Mesulam, & Nobre, 2005; Downing, 1988; Posner et al., 1980; Yu & Dayan, 2005b). More recently it has been shown that, in complex tasks, participants implicitly learn which visual signals provide task-relevant information, such as predicting which stimuli are likely to be presented, and that this information can be used to optimize performance in the task (Chun, 2000; Eckstein, Abbey, Pham, & Shimozaki, 2004).

As well as enhancing perceptual performance, expectations can also influence 'what' is perceived. Specifically, recent studies have shown that rapidly learned expectations can help determine the perception of bistable stimuli (Haijiang, Saunders, Stone, & Backus, 2006; Sterzer et al., 2008). Perception of such bistable stimuli is unstable, undergoing frequent reversals (van Ee, 2005) whose dynamics can be altered voluntarily by the observer (van Ee, van Dam, & Brouwer, 2005). In contrast, perception of simple stimuli is typically unambiguous and, seemingly, not so easily changed. Therefore, whether expectations can also alter the perception of simple stimuli that are not bistable is unclear.

A growing body of work suggests that perception is akin to Bayesian Inference (Knill & Pouget, 2004; Weiss, Simoncelli, & Adelson, 2002), where the brain represents sensory information probabilistically, in the form of probability distributions. Here it is assumed that in situations of uncer-

tainty, sensory information is combined with prior knowledge about the statistics of the world, serving to bias perception towards what is expected. This framework has been used to understand a great number of perceptual phenomena, such as why moving images appear to be moving slower when they are presented at low contrast (A. A. Stocker & Simoncelli, 2006), and the illusory ‘filling-in’ of discontinuous contours (Komatsu, 2006; Lee & Mumford, 2003), adding support to the idea that expectations can alter the appearance of simple unambiguous visual stimuli. However, in these studies, participants’ expectations (i.e. priors) are usually assumed to be acquired over long periods of time, through development and life experience. On the other hand, in the field of sensorimotor learning, it has been shown that participants can learn priors about novel statistics introduced during a psychophysical task, and that they combine this with information about their sensorimotor uncertainty in a manner that is consistent with a Bayes optimal process (Faisal & Wolpert, 2009; Körding & Wolpert, 2004). In the visual domain, how new sensory priors are learned is an open question.

Here we sought to understand whether stimulus expectations can be implicitly acquired through fast statistical learning, and if so, how such expectations are combined with visual signals to modulate perception of simple unambiguous stimuli. We examined this in the context of motion perception in a design where some motion directions were more likely to appear than others. Our hypothesis was that participants would automatically learn which directions were most likely to be presented and that these learned expectations would bias their perception of motion direction. A secondary hypothesis was that participants would solve the task using a Bayesian strategy, combining a learned prior of the stimulus statistics (the expectation) with their sensory evidence (the actual stimulus) in a probabilistic way.

## Methods

### Observers and stimuli

Twenty naive observers with normal or corrected-to-normal vision, participated in this experiment. All participants in the study gave informed written consent, received compensation for their participation and were recruited from the Riverside, CA area. The University of California, Riverside Institutional Review Board approved the methods used in the study, which was conducted in accordance with the Declaration of Helsinki.

Visual stimuli were generated using the Matlab programming language and displayed using Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) on Viewsonic P95f monitor running at 1024X768 at 100hz. The display luminance of the CRT monitor was made linear by means of an 8-bit lookup table. Participants viewed the display in a darkened room at a viewing distance of 100 cm with their motion constrained by a chin rest. Motion stimuli consisted of

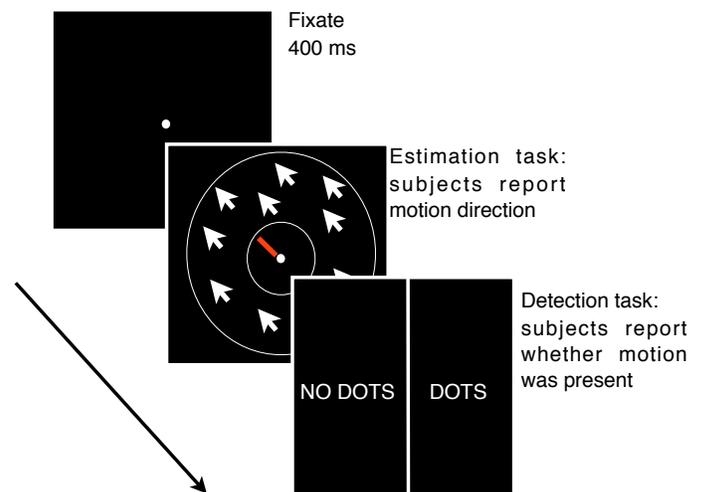


Figure 1: Sequence of events in a single trial. Each trial began with a fixation point, followed by the appearance of a motion stimulus. A central bar projecting from the fixation point was presented simultaneously with the motion stimulus, and allowed participants to estimate the direction of motion. After either participants had made an estimation, or a period of 3000 ms had elapsed, the stimulus disappeared and was replaced by a vertical line, with text to either side. Participants moved a cursor to either side of the line to indicate whether they had perceived the motion stimulus.

a field of dots (density: 2 dots/deg<sup>2</sup> at 100Hz refresh rate) moving coherently at a speed of 9°/sec within a circular annulus, with minimum and maximum diameter of 2.2° and 7° respectively. The background luminance of the display was set to 5.2 cd/m<sup>2</sup>.

### Procedure

At the beginning of each trial a central fixation point (0.5° diameter, 12.2cd/m<sup>2</sup>) was presented for 400 ms. With the fixation point still onscreen, the motion stimulus was then presented, along with a red bar which projected out (initial angle of bar randomized for each trial) from the fixation point (figure 1). The bar was located entirely within the center of the annulus containing the moving dots (length 1.1°, width 0.03°, luminance 3.4cd/m<sup>2</sup>). Participants indicated the direction of motion by orienting the red bar with a mouse, clicking the mouse button when they had made their estimate (estimation task). The display cleared when either the participant had clicked on the mouse, or a period of 3000 ms had elapsed. On trials where no motion stimulus was presented, the red bar still appeared and participants were required to estimate the perceived direction of motion as normal. Participants were instructed to fixate on the central point throughout this period. Participants’ reaction time in the estimation task determined how long the stimulus was presented for. On average this was equal to **1978±85ms** (standard error on the mean; see supplementary figure 7 for a plot of reaction time versus presented

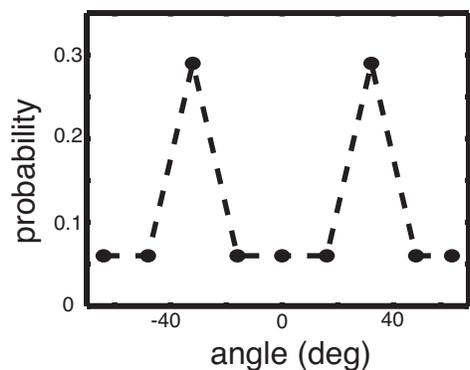


Figure 2: Probability distribution of presented motion directions. Two directions, 64° apart from each other, were presented in a larger number of trials than other directions. Motion direction is plotted relative to a reference direction at 0°, which was different for each subject.

motion direction). After the estimation task had finished, there was a 200ms delay before a vertical white line was presented at the center of the screen, with text to either side (reading ‘NO DOTS’ and ‘DOTS’ respectively). Participants moved a cursor to the right or left of this line to indicate whether they had or had not seen a motion stimulus (detection), and clicked the mouse button to indicate their choice. The cursor flashed green or red for a correct or incorrect detection response, respectively. The screen was then cleared and there was a 400 ms blank period before the beginning of the next trial.

Every 20 trials, participants were presented block feedback on the estimation task, with text displayed on screen telling participants what their average estimation error was in the previous 20 trials (e.g. “In the last 20 trials, your average estimation error was: 20°”). Block feedback, rather than trial by trial feedback was given, because we wanted to encourage participants to do their best at the estimation task, without interfering with their estimation behaviour (and biases) on each trial.

## Design

Participants took part in 2 experimental sessions lasting around 1 hour each, taken over successive days. Each session was divided into 5 blocks of 170 trials where all stimulus configurations were presented, making 1700 trials in total (850 trials per session).

Participants were presented stimuli at 4 different randomly interleaved contrast levels. The highest contrast level was at 1.7cd/m<sup>2</sup> above the 5.2cd/m<sup>2</sup> background. For each session there were 250 trials at zero contrast and 100 trials at high contrast. Contrasts of other stimuli were determined using 4/1 and 2/1 staircases on detection performance (García-Pérez, 1998). For each session there were 135 trials with the 2/1 staircase and 365 trials with the 4/1 staircase.

For the two staircased contrast levels, on a given trial the direction of motion could be 0° ±16°, ±32°, ±48° or

±64°, with respect to a central reference angle. To reduce potential biases in the population, we averaged results due to reference repulsion from cardinal motion directions (Rauber & Treue, 1998), this central motion direction was randomized across participants. We manipulated participants’ expectations about which motion directions were most likely to occur by presenting stimuli moving at ±32° more frequently than the others (figure 2). Therefore, at the 4/1 staircased contrast level, there were 130 trials per session with motion at -32° and +32°, and 15 trials per session for each of the other directions of motion. At the 2/1 staircased contrast level there were an equal number of stimuli moving in each of the predetermined directions: 15 trials per session for each motion direction. At the highest contrast level there were 25 trials per session with motion at -32° and +32° and 50 trials per session at completely random directions (among all possible directions, not just the predetermined directions used in the rest of the experiment).

## Data analysis

In the analysis of the estimation task, we looked only at trials where participants both reported seeing a stimulus and clicked on the mouse during stimulus presentation to indicate their estimate of motion direction. The first 100 trials from each session (~25 trials from each contrast staircase) were excluded from the analysis to allow the staircases to converge on stable contrast levels (supplementary figure 2a). Data was analyzed for the 12 (of 20) participants who could adequately perform both tasks according to our predetermined performance criteria of detection greater than 80% (quantified as the fraction of trials where participants both detect the stimulus and click on the mouse during stimulus presentation to estimate its direction) and mean absolute estimation error less than 30° with the highest contrast stimuli in both experimental sessions (supplementary figure 1; see supplementary materials for details of different participants’ performance). Importantly, our analysis of participants performance in the estimation task looked only at their responses to staircased contrast levels, and not their responses to the highest contrast stimuli, which we used to determine which participants should be included.

In the estimation task the variance of participants’ motion direction estimates tended to be quite large and varied greatly across different participants and motion directions. We hypothesized that this was due to the fact that in some trials participants made completely random estimates. Thus, data was fitted to the distribution:  $(1 - a)V(\mu, \kappa) + a/2\pi$ , where ‘a’ is the proportion of trials where the participant make random estimates, and  $V(\mu, \kappa)$  is a von Mises (circular normal) distribution with mean ‘μ’ and width determined by ‘1/κ’, given by:  $V(\mu, \kappa) = \exp(\kappa \cos(\theta - \mu)) / (2\pi I_0(\kappa))$ . Parameters were chosen by maximizing the likelihood of generating the data from the distribution. Participants’ estimation mean and standard deviation were taken as the circular mean and

standard deviation of the von Mises distribution,  $V(\mu, \kappa)$ . The average biases obtained using this method were qualitatively similar to those obtained through calculating the estimated direction by simply averaging over trials, while the variances were significantly smaller and with more consistency across participants and motion directions when the parametric fits were used. Therefore, in all of the following analysis we used this parametric method to quantify performance in the estimation task.

There was no significant interaction between experimental session and motion direction on the estimation bias or standard deviation ( $p = 0.11$  and  $p = 0.41$  respectively, 4-way within-subjects ANOVA). Therefore, we collapsed data across the two experimental sessions.

There was a considerable degree of overlap between the luminance levels achieved using both staircases. After discounting the first 100 trials from each session, the population averaged standard deviation in the luminance of the 2/1 and the 4/1 staircased levels over the course of one experimental session was  $0.051 \pm 0.001 \text{cd/m}^2$  and  $0.054 \pm 0.001 \text{cd/m}^2$  respectively; similar to the average luminance difference between the two levels ( $0.052 \pm 0.004 \text{cd/m}^2$ ). Further, there was no significant difference between the luminance levels achieved for both staircases ( $p = 0.23$ , 3-way within-subjects ANOVA). This was reflected in the estimation data: there was no significant difference between participants' estimation standard deviations for both staircased contrast levels ( $p = 0.12$ , 4-way within-subjects ANOVA). Therefore, we collapsed data across these contrast levels for all of the analysis described in the main text. Later, we looked at the effect of contrast level on participants' behaviour by separating participants' responses at different luminance levels, depending on their detection performance at different luminance levels. Details of this procedure are described in the supplementary materials.

To analyze the distribution of estimations when no stimulus was present, we constructed histograms of participants' responses, binned into  $16^\circ$  windows. We converted these response histograms into probability distributions, by normalizing them over all motion directions for each participant individually. There was no significant interaction between experimental session and motion direction on the response histograms ( $p = 0.87$ , 4-way within-subjects ANOVA). There was also no significant 3-way interaction between motion direction, experimental session and detection-response ( $p = 0.81$ , 4-way within-subjects ANOVA). Therefore we collapsed data across experimental sessions for analysis of the participants' responses when no stimulus was present.

In this study we were interested in how the uneven distribution of presented motion directions influenced participants' perception of the motion stimuli. By design, the probability distribution of presented motion stimuli was symmetrical around a central motion angle (figure 2). Therefore, we figured that any asymmetry in participants' estimation and detection behaviour for stimuli moving to

either side of the central motion direction was likely due to factors other than the distribution of presented stimuli that was used, such as participants' implicit biases, or 'reference biases' away from caudal motion directions (Rauber & Treue, 1998). To reduce the effect of such asymmetries from our analysis, and to increase the number of data points that were available for each experimental condition, we averaged data from points corresponding to when the presented motion stimuli was moving to either side of the central motion direction. For the estimation task, this also required reversing the sign of the estimation biases for stimuli moving anti-clockwise from the central motion direction before averaging (for 'unfolded' versions of figures 3a, 4a and 5 see supplementary figures 4 & 5).

## Results

### Effect of expectations on motion direction estimates when no stimulus present

First, we investigated whether participants learned to expect the most frequently presented motion directions. To assess this, we examined participants' estimation performance on trials where no stimulus was presented, but where they reported seeing a stimulus in the detection task, as well as clicking on the mouse to estimate its direction. On average this occurred on  $46 \pm 3$  trials for each participant ( $10.8 \pm 2\%$  of the total number of trials where no stimulus was presented). For this subset of trials, participants' estimation response probability varied significantly with motion direction, with a clear peak close to the most frequently presented motion directions ( $\pm 32^\circ$ ;  $p < 0.001$ , 3-way within-subjects ANOVA; figure 3a, grey). We quantified the probability ratio that participants made estimates that were close to the most frequently presented motion directions, relative to other directions, by multiplying the probability that they estimated within  $8^\circ$  of these motion directions by the total number of  $16^\circ$  bins ( $p_{rel} = p(\theta_{est} = \pm 32(\pm 8)^\circ) \cdot N_{bins}$ ). This probability ratio would be equal to 1 if participants were equally likely to estimate within  $8^\circ$  of  $\pm 32^\circ$  as they were to estimate within other  $16^\circ$  bins. We found that the median value of  $p_{rel}$  was significantly greater than 1, indicating that participants were strongly biased to report motion in the most frequently presented directions when no stimulus was presented (median( $p_{rel}$ ) = 2.7;  $p = 0.005$ , signed rank test, comparing  $p_{rel}$  to 1; figure 3b).

As on a large proportion of trials, the presented motion stimuli were moving in one of two directions, it is possible that participants could have habituated to automatically move the estimation bar towards one of these two directions, irrespective of their response in the detection task (note that the initial bar position was randomized on each trial and thus biases can't arise from just leaving the mouse in its initial location). In this case we would also expect their 'no-stimulus' estimation distributions to be biased towards the two most frequently presented directions for

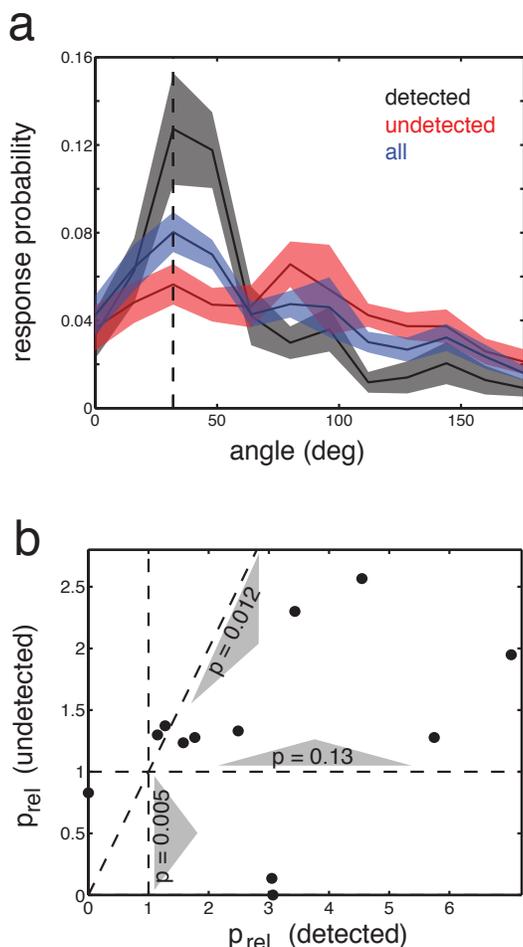


Figure 3: Estimation responses in the absence of a stimulus. (a) Probability distribution of participants' estimates of motion direction when no stimulus was present. Response distributions are plotted for all trials (blue), as well as the subset of trials where participants reported detecting a stimulus (grey) and trials where they didn't (red). Data points from either side of the central motion direction have been averaged together in this plot, so that the furthest left data point corresponds to the central motion direction, and the vertical dashed line corresponds to the most frequently presented motion directions ( $\pm 32^\circ$ ). Results are averaged over all participants and error bars represent within-subject standard error. (b) Probability ratio ( $p_{rel}$ ) that individual participants estimated within  $8^\circ$  from the most frequently presented motion directions ( $\pm 32^\circ$ ) relative to other  $16^\circ$  bins, plotted for trials where the stimulus was undetected versus trials where the stimulus was detected.  $p_{rel}$  was significantly greater than 1 for trials where participants reported detecting stimuli ( $p = 0.005$ , signed rank test), but was only marginally so when subjects failed to detect the stimulus ( $p=0.13$ ). Participants were also significantly more likely to estimate in the direction of the frequently presented motion directions on trials where they reported detecting stimuli, versus trials where they did not ( $p = 0.012$ ).

trials where they did not detect a stimulus. However, on trials where participants did not report seeing a stimulus in the detection task (but where they did click the mouse while

the stimulus was present to estimate its motion direction; on average this occurred on  $134 \pm 9$  trials for each participant;  $32 \pm 7\%$  of the total number of trials where no stimulus was presented), there was no significant variation in the estimation response probability with motion direction ( $p = 0.12$ , 3-way within-subjects ANOVA; figure 3a, red). Further, for these trials, participants were not significantly more likely to estimate close to the most frequently presented motion directions than other motion directions ( $\text{median}(p_{rel}) = 1.28$ ;  $p=0.13$ , signed rank test, comparing  $p_{rel}$  to 1; figure 3b). Indeed they were significantly more likely to report motion in the most frequently presented motion directions when they also reported detecting a stimulus, compared to when they did not ( $p = 0.012$ , signed rank test, comparing the values of  $p_{rel}$  obtained for trials where participants either did or did not report seeing a stimulus in the detection task; figure 3b).

It could be argued that we would observe similar results if participants' expectations influenced their behaviour in the detection task, but not in the estimation task. Thus, in the absence of a presented stimulus, they would be more likely to report detecting a stimulus when they mistakenly perceived motion in one of the two most frequently presented motion directions, although their estimation responses would be unaltered by their expectations. In this case participants' estimation responses would be distributed uniformly when we looked at data from all trials where no stimulus was presented (regardless of their response in the detection task). This was not what we found: when we looked at data from all zero-stimulus trials, participants' estimation response probability varied significantly with motion direction ( $p < 0.001$ , 3-way within-subjects ANOVA; figure 3a, blue) and they were biased to report motion in the two most frequently presented directions ( $\text{median}(p_{rel}) = 1.71$ ;  $p < 0.001$ , signed rank test comparing  $p_{rel}$  to 1). However, the size of this bias was reduced, compared to the case when we looked only at trials where participants detected stimuli ( $p = 0.027$ , signed rank test comparing the values of  $p_{rel}$  obtained for all trials with trials where participants reported seeing a stimulus in the detection task).

Another response strategy that could have produced similar results is if, when participants were uncertain about the stimulus motion direction, they made estimations that were influenced by the stimulus presented immediately beforehand. In this case, we would expect the observed biases in participants' no-stimulus estimation distributions to disappear when we excluded trials that were immediately preceded by stimuli moving in the most frequently presented directions ( $\pm 32^\circ$ ). However, when we excluded these trials from our analysis, participants' zero-stimulus estimations (for trials where they reported detecting a stimulus) were still strongly biased towards the two most frequently presented directions ( $\text{median}(p_{rel})=2.11$ ;  $p=0.026$ , signed rank test, comparing  $p_{rel}$  to 1).

Taken together, our results indicate that the zero-stimulus biases we observed were not due to 'response strategies', but rather, were perceptual in origin: partici-

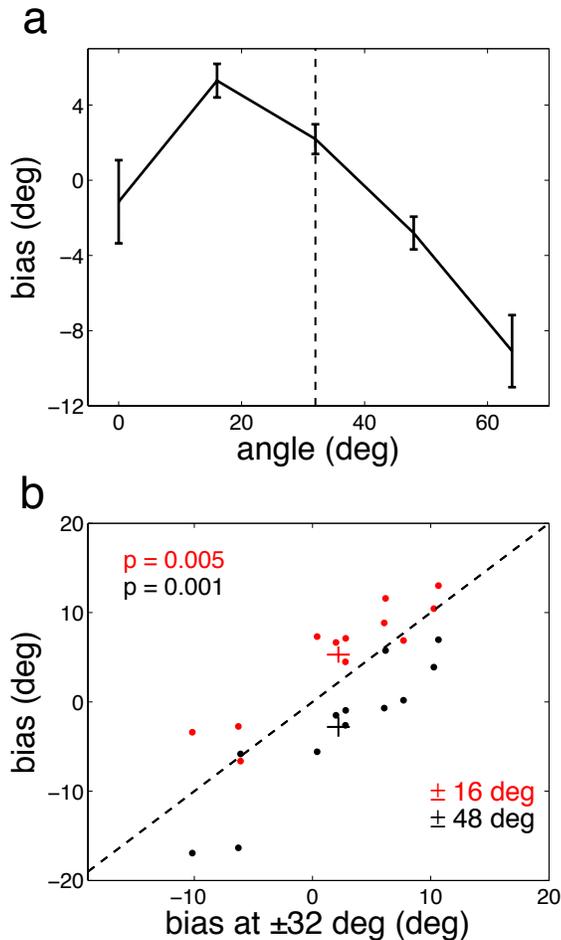


Figure 4: Effect of expectations on estimation biases. (a) Participants' mean estimation bias is plotted against presented motion direction. Data points from either side of the central motion direction have been averaged together, so that the furthest left point corresponds to the central motion direction, and the vertical dashed line corresponds to data taken from the two most frequently presented motion directions ( $\pm 32^\circ$ ). Results are averaged over all participants and error bars represent within-subject standard error. (b) The estimation bias for stimuli moving at  $\pm 48^\circ$  (black) and  $\pm 16^\circ$  (red) from the central motion direction, plotted against the estimation bias at  $\pm 32^\circ$ , for each participant. Again, data from stimuli moving to both sides of the central motion direction has been averaged together, with the sign of the bias for stimuli moving anti-clockwise from the central motion direction (i.e.  $-48^\circ$ ,  $-32^\circ$  and  $-16^\circ$ ) reversed before averaging. The red and black crosses mark the population mean of both distributions, with the length of the lines on the crosses equal to the standard error.

pants 'hallucinated' motion in the most frequently presented directions when no stimulus was displayed. Further, these hallucinations developed extremely quickly. On trials where no stimulus was presented, but where participants reported detecting a stimulus, they were significantly more likely to estimate within  $8^\circ$  of  $\pm 32^\circ$ , than other directions, after a period of only 200 trials ( $p = 0.008$ , signed rank test,

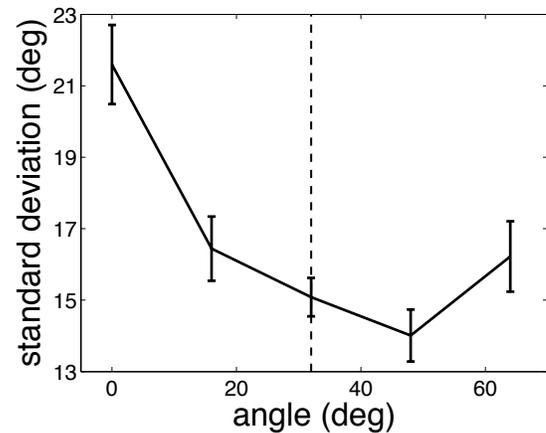


Figure 5: Effect of expectations on the standard deviation of estimations. The standard deviation in participants' estimation distributions is plotted against presented motion direction. Data points from either side of the central motion direction have been averaged together, so that the furthest left point corresponds to the central motion direction, and the vertical dashed line corresponds to data taken from the two most frequently presented motion directions ( $\pm 32^\circ$ ). Results are averaged over all participants and error bars represent within-subject standard error.

comparing  $p_{rel}$  to 1 after 200 trials; see supplementary figure 3), indicating rapid learning of motion direction expectations.

### Effect of expectations on motion direction estimates when stimulus was presented

We next asked whether these learned expectations would bias participants' perceptions of real motion stimuli. Figure 4a shows the population averaged estimation bias, plotted against motion direction. In this plot, data points corresponding to presented stimuli moving to either side of the central motion direction have been averaged together (making sure to reverse the sign of the estimation bias when the presented stimuli was anti-clockwise from the central motion direction before averaging; see supplementary figure 4 for an alternative version of this plot without averaging across the central motion direction). In this plot the curve has a negative slope around  $+32^\circ$ , which itself was unbiased. This indicates that estimations were attractively biased towards stimuli moving at  $+32^\circ$  (and by symmetry, also to motion at  $-32^\circ$ ). Estimates of the central motion direction were unbiased, while estimates at  $+16^\circ$  were positively biased, away from the centre and towards stimuli moving at  $+32^\circ$  (again, by symmetry, stimuli moving at  $-16^\circ$  were biased away from the centre, towards stimuli moving at  $-32^\circ$ ). Note that the apparent asymmetry in figure 4a is expected, and is due to the fact that the data points at  $0^\circ$  and  $64^\circ$  are not equivalent:  $0^\circ$  lies midway between the two most frequently presented directions, while  $+64^\circ$  is on the edge of the distribution of presented motion directions (see figure 2).

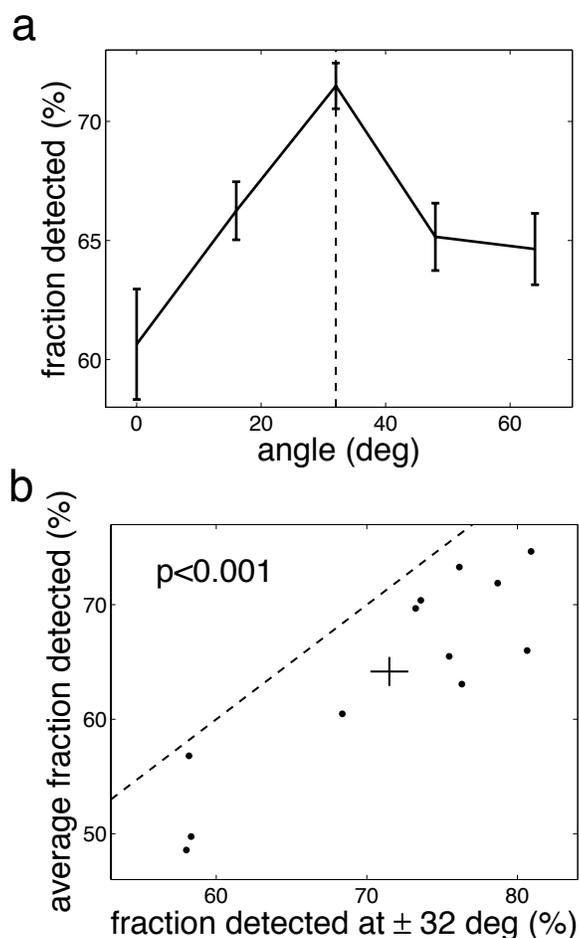


Figure 6: Effect of expectations on detection performance. (a) The fraction of trials where participants correctly detected a motion stimulus is plotted against presented motion direction. Data points from either side of the central motion direction have been averaged together, so that the furthest left point corresponds to the central motion direction, and the vertical dashed line corresponds to data taken from the two most frequently presented motion directions ( $\pm 32^\circ$ ). Results are averaged over all participants and error bars represent within-subject standard error. (b) The fraction of trials where participants correctly detected a stimulus, averaged over all presented motion directions except for  $\pm 32^\circ$ , plotted against the fraction of trials where participants correctly detected a stimulus moving at  $\pm 32^\circ$ , for each participant. The black cross marks the population mean, with the length of the lines on the cross equal to the standard error.

Overall, there was a significant effect of motion direction on the estimation bias ( $p < 0.001$ , 3-way within-subjects ANOVA).

We wanted to quantify the extent to which individual participants' estimates were biased towards the most frequently presented motion directions. For a participants whose estimates were attractively biased towards stimuli moving at  $+32^\circ$ , we would expect their estimates of stimuli moving at  $+48^\circ$  and  $+16^\circ$  to be positively and negatively bi-

ased respectively, compared to their estimation bias for stimuli moving at  $+32^\circ$  (and by symmetry, we would also expect the converse to hold for stimuli moving anti-clockwise from the central direction: for a participant whose estimates were attractively biased towards stimuli moving at  $-32^\circ$ , we would expect the bias at  $-48^\circ$  and  $-16^\circ$  to be negatively biased and positively biased respectively, compared to their estimation bias for stimuli moving at  $-32^\circ$ ). Figure 4b plots individual participants' estimation bias for stimuli moving at  $\pm 48^\circ$  and  $\pm 16^\circ$  versus their estimation bias at  $\pm 32^\circ$  (plotted in black and red respectively). Note that, similarly to figure 4a, we averaged data from motion directions moving to either side of the central motion directions in this plot, making sure to reverse the sign of the bias for stimuli moving anti-clockwise from the central motion direction. After doing this, the computed estimation biases at  $\pm 48^\circ$  and  $\pm 16^\circ$  were significantly smaller and larger respectively than the bias at  $\pm 32^\circ$  ( $p = 0.005$  and  $p = 0.001$  respectively, signed rank test). This indicates that on average, participants were biased to estimate stimuli as moving in directions that were closer to the most frequently presented motion directions ( $\pm 32^\circ$ ) than they actually were.

Stimuli in-between  $\pm 32^\circ$  were expected to be biased by both frequently presented directions and thus we expected that these directions should yield larger standard deviations in estimated angles than those outside of this range. Figure 5 plots the population averaged standard deviation of estimations against motion direction. Again, for this plot, data points from either side of the central motion direction have been averaged together. The estimation standard deviation was greatest for the central motion direction at  $0^\circ$ , and smallest for motion directions that were closer to the most frequently presented directions ( $\pm 16^\circ$ ,  $\pm 32^\circ$  and  $\pm 48^\circ$ ). As with the estimation biases, there was a significant effect of motion direction on the estimation standard deviation ( $p < 0.001$ , 3-way within-subjects ANOVA).

### Effect of expectations on detection performance and reaction time

One of our interests was the extent to which stimulus expectations influenced participants' performance in the detection task. To test this, we measured the fraction of trials where participants both detected stimuli and clicked on the mouse during stimulus presentation, as a function of motion direction (figure 6a). Participants were significantly more likely to detect stimuli moving in the most frequently presented motion directions ( $71.5 \pm 2.5\%$  detected at  $\pm 32^\circ$  versus  $64.2 \pm 2.5\%$  detected over all other motion directions;  $p < 0.001$  signed-rank test; figure 6b). Overall, there was a significant effect of motion direction on the fraction detected ( $p = 0.002$ , 3-way within-subjects ANOVA).

Another measure that could reflect how easily participants detected stimuli was their reaction time in clicking the mouse during stimulus presentation. For trials where they detected a stimulus, participants' reaction time was significantly reduced for the most frequently presented mo-

tion directions, relative to other motion directions ( $1924 \pm 86$ ms at  $\pm 32^\circ$  versus  $1991 \pm 85$ ms over all other motion directions;  $p < 0.001$ , signed rank test; supplementary figure 7). Overall, there was a significant effect of motion direction on participants' reaction time ( $p = 0.003$ , 3-way within-subjects ANOVA).

## Modeling

To understand the nature of the biases in motion direction estimation that we observed, we tested among alternative models of how participants' expectations may be combined with the presented stimulus to produce the observed response distributions. Two classes of models were considered. The first class of model assumed that participants developed response strategies unrelated to perceptual changes. The second class of model assumed that participants solved the task using a Bayesian strategy, combining a learned prior of the stimulus statistics (the expectation) with their sensory evidence (the actual stimulus) in a probabilistic way. These models simulate the estimation distributions in the case where participants judged the stimulus to be present.

### Multiple-strategy 'response bias' models

The first two models looked at whether participants' behaviour could be attributed to a 'response bias'. The key assumption in both of these models was that participants followed different strategies on different trials: for example, by making an unbiased estimate of motion direction on a fraction of the trials, and by estimating one of the most frequently presented motion directions on other trials.

The first model ('ADD1') assumed that when participants were unsure about which motion direction they had perceived they made an estimate that was close to one of the two most frequently presented motion directions.

In this model, on each trial, participants make a sensory observation of the stimulus motion direction,  $\theta_{obs}$ . We parameterize the probability of observing the stimulus to be moving in a direction  $\theta_{obs}$  by a von Mises (circular normal) distribution centered on the actual stimulus direction and with width determined by  $1/\kappa_l$ :

$$p_l(\theta_{obs}|\theta) = V(\theta, \kappa_l) \quad (1)$$

On most trials we assume that participants make a perceptual estimate of the stimulus motion direction ( $\theta_{perc}$ ) that is based entirely on their sensory observation so that,  $\theta_{perc} = \theta_{obs}$ . However, on a certain proportion of trials, when participants are uncertain about whether a stimulus was present or not, they resort to their 'expectations', by making a perceptual estimate that is sampled from a learned distribution,  $p_{exp}(\theta)$ . For simplicity, we parameterize this distribution as the sum of two circular normal distributions, each with width determined by  $1/\kappa_{exp}$ , and centered on motion directions  $-\theta_{exp}$  and  $\theta_{exp}$  respectively:

$$p_{exp}(\theta) = \frac{1}{2}[V(-\theta_{exp}, \kappa_{exp}) + V(\theta_{exp}, \kappa_{exp})] \quad (2)$$

Finally, we accommodate for the fact that there will be a certain amount of noise associated with moving the 'estimation bar' to indicate which direction the stimulus is moving in, as well as allowing for a fraction of trials ' $\alpha$ ', where participants make estimates that are completely random. Thus, the estimation response  $\theta_{est}$  is related to the perceptual estimate  $\theta_{perc}$  via the equation:

$$p(\theta_{est}|\theta_{perc}) = (1 - \alpha)V(\theta_{perc}, \kappa_m) + \alpha \quad (3)$$

Bringing all this together, the distribution of estimation responses for a single participant is given by:

$$p(\theta_{est}|\theta) = (1 - \alpha)[(1 - a(\theta)) \cdot p_l(\theta_{obs} = \theta_{est}|\theta) + a(\theta) \cdot p_{exp}(\theta_{est})] * V(0, \kappa_m) + \alpha \quad (4)$$

where the asterisk denotes a convolution and  $a(\theta)$  determines the proportion of trials that participants sampled from the 'expected' distribution,  $p_{exp}(\theta)$ . For this model, free parameters that were fitted to the estimation data for each participant were the centre and width of participants' 'expected' distributions (determined by  $\theta_{exp}$  &  $1/\kappa_{exp}$  respectively), the width of their sensory likelihood (determined by  $1/\kappa_l$ ), the fraction of trials where they made estimates by sampling from their 'expected' distribution ( $a(\theta)$ ), the magnitude of the 'motor' noise in their responses (determined by  $1/\kappa_m$ ) and the fraction of trials where they made estimations that were completely random ( $\alpha$ ).

The second 'response-bias' model ('ADD2') assumed a more complex strategy, such that when participants were unsure of stimulus direction, they made estimates that were preferentially sampled from different proportions of their 'expected' distribution. Crucially, the portion of this 'expected' distribution that was sampled from depended on the actual stimulus motion direction.

Here, the expected distribution  $p_{exp}(\theta)$  was divided into two parts:

$$p_{anti-clockwise}(\theta) = V(-\theta_{exp}, \kappa_{exp}) \quad (5)$$

$$p_{clockwise}(\theta) = V(\theta_{exp}, \kappa_{exp}) \quad (6)$$

As before, on a single trial, participants made estimates that were either equal to their sensory observation  $\theta_{obs}$ , or sampled from a learned distribution of expected motion directions. However, instead of sampling from a single distribution of expected motion directions,  $p_{exp}(\theta)$ , participants could now make estimates that were sampled either from the distributions  $p_{anti-clockwise}(\theta)$  or  $p_{clockwise}(\theta)$ , with a probability that was dependent on the actual stimulus motion direction. For example, on a single trial, a participant might be aware that the stimulus was moving 'clockwise from centre', and thus would be more likely to make an estimate that was sampled from the distribution,  $p_{anti-clockwise}(\theta)$ , than from  $p_{clockwise}(\theta)$ .

This more complex response strategy results in a distribution of estimation responses given by:

$$\begin{aligned}
p(\theta_{est}|\theta) = & (1 - \alpha)[(1 - a(\theta) - b(\theta)) \cdot p_l(\theta_{obs} = \theta_{est}|\theta) \\
& + a(\theta) \cdot p_{anti-clockwise}(\theta_{est}) \\
& + b(\theta) \cdot p_{clockwise}(\theta_{est})] * V(0, \kappa_m) + \alpha
\end{aligned} \quad (7)$$

where ‘ $a(\theta)$ ’ and ‘ $b(\theta)$ ’ were additional free parameters that determined the proportion of trials where participants sampled from each distribution.

Finally, we considered variations to the *ADD1* and *ADD2* models (denoted ‘*ADD1\_mode*’ and ‘*ADD2\_mode*’ respectively) where, on trials where participants were unsure of the stimulus motion direction, they made perceptual estimates that were equal to the mode of the ‘expected’ distribution. These models are equivalent to the *ADD1* and *ADD2* models, with ‘ $1/\kappa_{exp}$ ’ set to zero.

### Bayesian model

The second class of models assumed that participants combined a learned prior of the stimulus directions with their sensory evidence in a probabilistic manner. Specifically, unlike the previous models, where on individual trials participants either rely entirely on their sensory observations or on their expectations, in the Bayesian model participants make estimations based on a combination of both their sensory observation and expectations. A schematic of this model class is shown in figure 7.

As before, we assume that on a single trial, participants make noisy sensory observations of the stimulus motion direction ( $\theta_{obs}$ ), with a probability  $p_l(\theta_{obs}|\theta) = V(\theta, \kappa_l)$ . From Bayes’ rule, the posterior probability that the stimulus is moving in a particular direction  $\theta$ , given a sensory observation  $\theta_{obs}$ , is obtained by multiplying the likelihood function ( $p_l(\theta_{obs}|\theta)$ ), with the prior probability ( $p_{prior}(\theta)$ ):

$$p(\theta|\theta_{obs}) \propto p_{prior}(\theta) \cdot p_l(\theta_{obs}|\theta) \quad (8)$$

While participants cannot access the ‘true’ prior,  $p_{prior}(\theta)$ , directly, we hypothesized that they learned an

approximation of this distribution, denoted ‘ $p_{exp}(\theta)$ ’. In our model this ‘learned prior’ was parameterized similarly to  $p_{exp}(\theta)$  in *ADD1* (see equation 2).

We assume that participants make perceptual estimates of motion direction,  $\theta_{perc}$ , by choosing the mean of the posterior distribution, so that :

$$\theta_{perc} = \frac{1}{Z} \int \theta \cdot p_{exp}(\theta) \cdot p_l(\theta_{obs}|\theta) d\theta \quad (9)$$

where ‘ $Z$ ’ is a normalization constant. An alternative choice would be for the perceptual estimate to be given by the maximum of the posterior distribution. For our work both methods gave qualitatively identical results.

We accounted for the ‘motor noise’ associated with making the estimation response in a similar way to the previous models. For this model, the free parameters that were fitted to the estimation data for each participant were the centre and width of participants’ ‘expected’ distribution (determined by  $\theta_{exp}$  &  $1/\kappa_{exp}$  respectively), the width of their sensory likelihood (determined by  $1/\kappa_l$ ), the magnitude of the ‘motor’ noise in their responses (determined by  $1/\kappa_m$ ) and the fraction of trials where they made estimations that were completely random ( $\alpha$ ). We included two variants of the Bayesian model: ‘*BAYES\_Lvar*’, where the width of the likelihood function was allowed to vary with the stimulus motion direction, and ‘*BAYES\_Lconst*’, where it was held constant.

### Inferring the parameters for each model

At the highest contrast, the stimulus was clearly visible, so we assumed that the perceptual uncertainty was close to zero ( $1/\kappa_l \rightarrow 0$ ). Therefore for all models, the distribution of estimations should be given by equation (3), with the substitution,  $\theta_{perc} = \theta$ . We used this equation to fit participants’ estimation distributions at high contrast (by maximizing the log probability of getting the observed the data; see later), thus allowing us to approximate the ‘motor

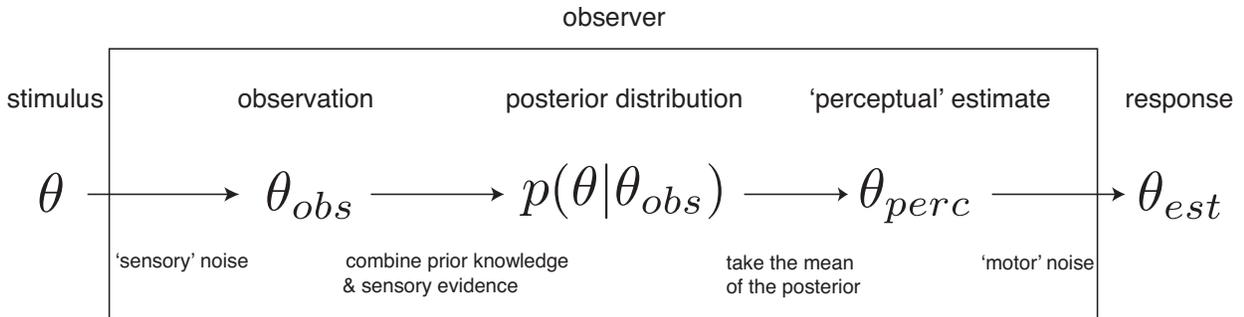


Figure 7: Bayesian model for estimation. The posterior distribution of possible stimulus motion directions is constructed by combining prior knowledge about likely motion directions (the expectation) with the available sensory evidence (based on a noisy observation,  $\theta_{obs}$ ) probabilistically. A perceptual estimate is made by taking the mean of the posterior distribution. This posterior distribution is used to make a perceptual estimate ( $\theta_{perc}$ ). Additional ‘motor noise’ is added to this perceptual estimate to produce the final estimation response ( $\theta_{est}$ )

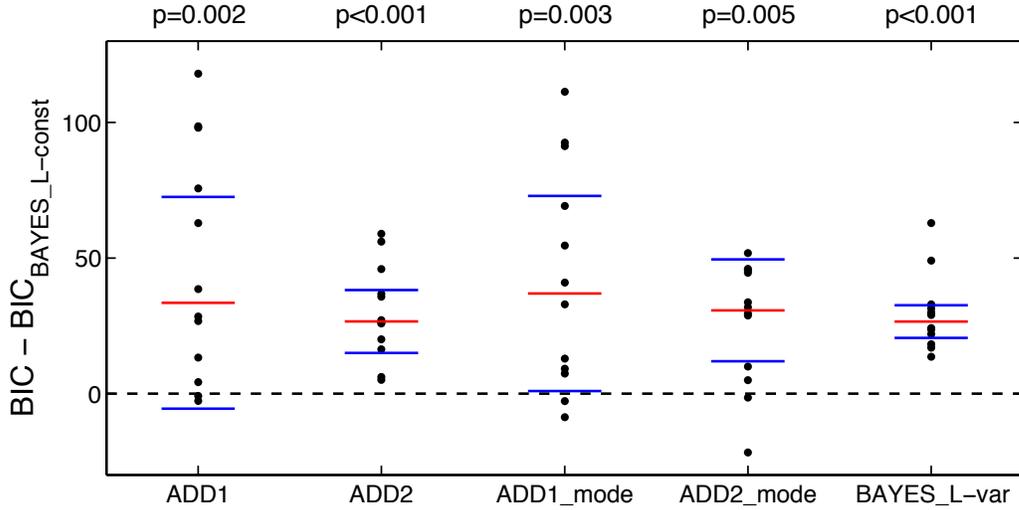


Figure 8: Model comparison. The Bayesian information criterion ( $BIC$ ) evaluated with each model, subtracted by the  $BIC$  evaluated with the  $BAYES\_L\text{-const}$  model, is plotted separately for each participant. Median values are indicated by horizontal red lines, 25<sup>th</sup> and 75<sup>th</sup> percentiles by horizontal blue lines. Values greater than zero indicate that the  $BAYES\_L\text{-const}$  model provided the best description of the data.  $p$ -values indicate whether the median ' $BIC - BIC_{BAYES\_L\text{-const}}$ ' was significantly different from zero for each model (signed rank test).

noise' (determined by  $1/\kappa_m$ ) for each participant.

As with the rest of our data analysis, we modelled participants responses to stimuli at both staircased contrast levels (although see supplementary materials). Also, as all three models looked only at the estimation task, effectively ignoring the detection response, we initially looked only at data where participants detected the motion stimulus (see supplementary materials for a version of the Bayesian model which incorporates the detection task).

For each model, and for a particular set of parameters ' $M$ ', we were able to calculate the probability of making an estimate ' $\theta_{est}$ ' given a stimulus moving in a direction ' $\theta$ ' ( $p(\theta_{est}|\theta; M)$ ). Assuming that participants' responses on each trial were independent, this allowed us to calculate the likelihood of generating our experimental data ' $D$ ' from the particular model and parameter set ' $M$ '. We then chose model parameters to fit the data for each participant by maximizing the log of the likelihood function:

$$M = \operatorname{argmax}_M \left[ \sum_i^{n_{\text{trials}}} \log(p(\theta_{est} = \theta_{i,\text{data}}|\theta_i)) \right] \quad (10)$$

where the summation was taken over all trials, and ' $\theta_i$ ' and ' $\theta_{i,\text{data}}$ ' represent the presented motion direction and the estimation response on the  $i^{\text{th}}$  trial respectively. We found the maximum of the likelihood function using a simplex algorithm (the Matlab function 'fminsearch'). We were concerned that for some participants our model fits might converge to local rather than global maxima. To reduce this possibility, we ran the model fits with a range of initial values for  $\kappa_l$  and  $\kappa_{exp}$  (' $1/\sqrt{\kappa_l}$ ' and ' $1/\sqrt{\kappa_{exp}}$ ' were varied independently in  $2^\circ$  increments, between  $1^\circ$  and  $21^\circ$ ), selecting the model fit that produced the highest value for the log-

likelihood. The results obtained were also found to be robust to changes in all of the other initial parameter values.

The models varied greatly with respect to the number of parameters that they required to fit the data. Excluding  $\kappa_m$  (as this was obtained from the high contrast responses, not the low contrast responses that were the principle area of investigation),  $ADD1$  and  $ADD2$  required 9 and 14 free parameters respectively:  $\kappa_l$ ,  $\theta_{exp}$ ,  $\kappa_{exp}$  and  $\alpha$ , plus 5 values for  $a(\theta)$  and, for  $ADD2$ , another 5 values for  $b(\theta)$  (one for each presented motion direction).  $ADD1\_mode$  and  $ADD2\_mode$  required 8 and 13 free parameters respectively (one less parameter than  $ADD1$  and  $ADD2$  respectively, as  $\kappa_{exp}$  was no longer a free parameter).  $BAYES\_L\text{-const}$  required only 4 free parameters ( $\kappa_l$ ,  $\theta_{exp}$ ,  $\kappa_{exp}$  and  $\alpha$ ).  $BAYES\_L\text{-var}$  required 8 free parameters (including a value for  $\kappa_l$  for each presented motion direction).

### Model comparison

We assessed how well each of the models accounted for the estimation distribution using a metric called the 'Bayesian information criterion' ( $BIC$ ), defined as:

$$BIC = -2 \cdot \ln(L) + k \cdot \ln(n) \quad (11)$$

where, ' $L$ ' is the likelihood of generating the experimental data from the model, ' $k$ ' is the number of parameters in the model and ' $n$ ' is the number of data points available. In general, given two estimated models, the model with the lower value of  $BIC$  is the one to be preferred (Schwarz, 1978). The first term of this expression accounts for the error between the data and the model predictions, while the second term represents a penalty for including too much complexity in the model.

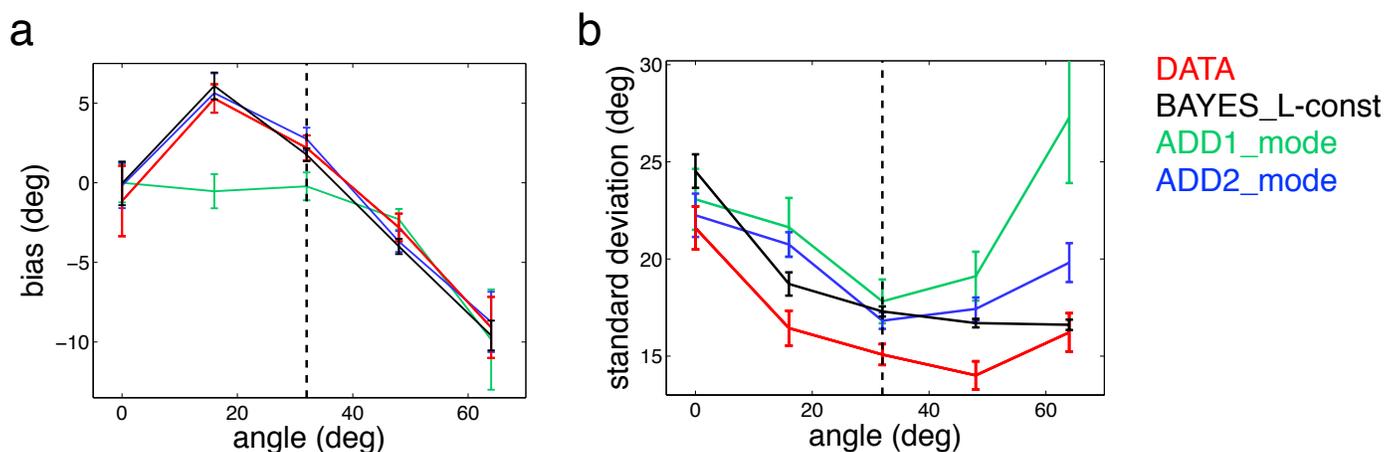


Figure 9: Predicted biases (a) and standard deviations (b) for each model. Predictions for the *ADD1-mode* model (green), the *ADD2-mode* model (blue) and the *BAYES\_L-const* model (black) are plotted alongside the experimental data (red). In both plots, data points from either side of the central motion direction have been averaged together, so that the furthest left point corresponds to the central motion direction, and the vertical dashed line corresponds to the most frequently presented motion directions. In all plots, results are averaged over all participants and error bars represent within-subject standard error.

Figure 8 plots, for each participant, the *BIC* obtained with each model, subtracted by the *BIC* obtained with the *BAYES\_L-const* model. From this plot we can see that the *BIC* values obtained with the *ADD1*, *ADD2*, *ADD1\_mode*, *ADD2\_mode* and *BAYES\_L-var* models were significantly greater than the *BIC* values obtained with the *BAYES\_L-const* model ( $p=0.002$ ,  $p<0.001$ ,  $p=0.003$ ,  $p=0.005$  and  $p<0.001$  respectively; signed-rank test). Thus, while a small minority of participants were not best fitted by the *BAYES\_L-const* model (2 participants exhibited a lower *BIC* value with the *ADD1* model, 2 participants exhibited a lower *BIC* value with the *ADD1\_mode* model and 2 participants exhibited a lower *BIC* value with the *ADD2\_mode* model), this model provided the best description of the data for the majority of participants.

Each of the models described attempted to fit the estimation distributions for each participant. To achieve a qualitative understanding of how the estimation distributions predicted by each of the models compared to the experimental data, we analyzed the predicted estimation biases and standard deviations. As the *ADD1\_mode* and the *ADD2\_mode*, and the *BAYES\_L-const* models provided better fits to the data than the other models, we only analyze here the predicted estimation biases and standard deviations for these three models. In our previous analysis of the experimental data, we parameterized participants' estimation distributions as the sum of a circular normal distribution and a 'flat' background probability (to account for the proportion of trials where they made random estimations). Participants estimation means and standard deviations were then taken as the centre and width of the fitted circular normal distribution respectively. To be consistent with this, we computed biases and standard deviations from the estimation distributions predicted by each model in an identical way.

Figure 9 shows the estimation biases and standard deviations predicted by each of the models, plotted alongside the experimental data. Both the *BAYES\_L-const* and *ADD2\_mode* models provided a good fit for the population averaged estimation biases (mean absolute error of  $0.75^\circ$ , and  $0.62^\circ$  for the *BAYES\_L-const* and *ADD2\_mode* models respectively). The *ADD1\_mode* model, however, was unable to reproduce the repulsive biases away from the central motion direction (at  $\pm 16^\circ$ ) that were observed experimentally (mean absolute error of  $2.14^\circ$ ; figure 9a). This was also reflected in the fits of individual participants' estimation biases (quantified by calculating the mean absolute error for the fits of the estimation biases separately for each participant, averaged over motion directions). The error in the fits of the individual participants' estimation biases was significantly smaller for the *BAYES\_L-const* model than for the *ADD1\_mode* model ( $p<0.001$ , signed rank test), while there was no significant difference between the *BAYES\_L-const* and *ADD2\_mode* models.

The fact that the *ADD1\_mode* model was unable to fit the experimentally observed repulsive biases away from the central motion direction can be explained by the fact that for this model we parameterized the 'expected' distribution of motion direction,  $p_{exp}(\theta)$ , to be symmetrical around  $0^\circ$ . Thus, even in the extreme case where all responses are sampled from this distribution, there would only be an attractive bias towards the central motion direction.

The *BAYES\_L-const* model produced estimation standard deviations that varied with motion direction in a qualitatively similar way to the experimental data, (with a maximum at  $0^\circ$ , decreasing for stimuli moving further from the central motion direction), although in general, the model predicted values that were slightly larger than what was observed experimentally (figure 9b). The fits for the estimation standard deviation produced by the *ADD1\_mode* and *ADD2\_mode* were worse than the *BAYES\_L-const* model

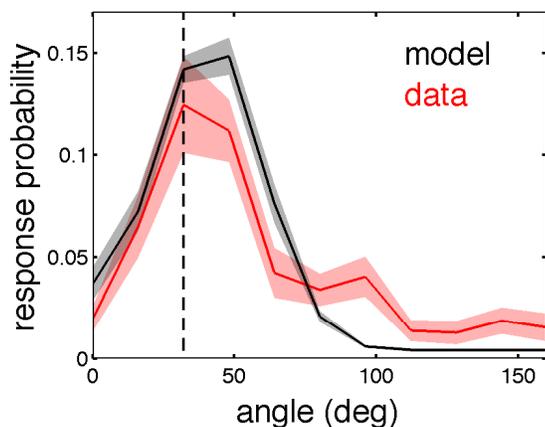


Figure 10: Predicted estimation response probability distributions for trials where no stimulus is presented, but where participants reported detecting a stimulus. Model predictions (grey; *BAYES\_dual* model, see supplementary materials for details) are plotted alongside the experimental results (red). Data points from either side of the central motion direction have been averaged together in this plot, so that the furthest left data point corresponds to the central motion direction, and the vertical dashed line corresponds to the most frequently presented motion directions ( $\pm 32^\circ$ ). Results are averaged over all participants and shaded error bars represent within-subject standard error.

(mean absolute error of  $5.11^\circ$ , and  $2.74^\circ$  for the *ADD1\_mode* and *ADD2\_mode* models respectively, compared to  $2.17^\circ$  for the *BAYES\_L-const* model), and did not vary with motion in a way that was similar to the experimental data. However, the error in the fits of the individual participants estimation standard deviations (quantified by calculating the mean absolute error for the fits of the estimation standard deviation separately for each participant, averaged over motion directions) was not significantly different between the models ( $p=0.91$  and  $p=0.34$  respectively for comparisons of the *ADD1\_mode* and *ADD2\_mode* models with the *BAYES\_L-const* model; signed rank test)

While all the free parameters in the *BAYES\_L-const* model ( $\sigma_l$ ,  $\theta_{exp}$ ,  $\sigma_{exp}$  and  $\alpha$ ) were held constant across presented motion directions, in order for the ‘response bias’ models (*ADD1*, *ADD2*, *ADD1\_mode* and *ADD2\_mode*) to fit the data, additional free parameters were required ( $a(\theta)$  and  $b(\theta)$ ) which had to be varied between different presented motion directions. Thus, for the *ADD1* and *ADD2* models to be valid, participants would have had to alter their response strategy, varying the proportion of trials where they sampled from their ‘expected’ probability distributions, depending on the direction of the presented stimulus. In addition, the *ADD1\_mode* and *ADD2\_mode* models assumed that when participants were unsure about the presented motion direction, they made a perceptual estimate of motion direction that was exactly the same on each trial. This seems unrealistic: in reality there would be some trial to trial variation in the expected motion direction.

In summary, *BAYES\_L-const* exhibited significantly smaller *BIC* values than all of the other models, as well as producing fits for the estimation biases and standard deviation that were at least as good as the response bias models, despite the fact that it had fewer free parameters (4 parameters, as opposed to 9, 14, 8, and 13 parameters for *ADD1* and *ADD2*, *ADD1\_mode* and *ADD2\_mode* respectively), leading us to conclude that it provided the best description of participants’ behavior. Overall, our results argue against the hypothesis that the observed estimation biases were produced by ‘response strategies’ unrelated to perceptual changes, but rather support the hypothesis that participants performed the task using a Bayesian strategy, where a learned a prior of expected stimulus directions was combined with their sensory evidence in a probabilistic way.

### Modeling estimation responses in the absence of a stimulus

We were interested to see whether the prior and likelihood distribution that we derived to fit participants’ response distributions when a stimulus was present were sufficient to explain their estimation performance in the absence of any stimulus.

While the original *BAYES\_L-const* model ignored the detection task, in order to analyze participants’ ‘no-stimulus’ behaviour it was important to incorporate this into our model. The full model, *BAYES\_dual*, which is of the same form as the original bayesian model, with the exception that it simulates the detection task, is described in the supplementary materials. The *BAYES\_dual* model required 3 additional parameters: participants’ prior expectation that a stimulus would be presented on each trial, the probability that participants made sensory observations of the stimulus as being present, on trials where a stimulus was presented, and on trials where no stimulus was presented (see supplementary materials). Importantly, these parameters were fitted using only data from trials where the stimulus was presented, and not zero-stimulus trials, which was what we were aiming to predict.

Figure 10 shows the estimation distributions predicted by this model for trials where there was no stimulus present, but where participants detected a stimulus (black), plotted alongside the experimentally measured distribution (red). The average ‘zero-stimulus’ estimation distribution predicted by the model, provided a good fit for the population averaged estimation distributions, with an  $R^2$  value of 0.71. The behaviour of individual participants was also well predicted by the model: the fits for participants’ zero stimulus estimation distributions had a positive  $R^2$  value for 8 out of 12 of them. For these participants, the median  $R^2$  value was 0.65 (0.46, 0.83; 25<sup>th</sup> & 75<sup>th</sup> percentiles). The fact that the majority of participants’ behaviour in the absence of a stimulus could be predicted, based solely on their estimation responses in the presence of a stimulus, provides strong evidence in favor of the Bayesian model put forward here.

## Discussion

We found that participants quickly and automatically developed expectations for the most frequently presented directions of motion. On trials where no stimulus was presented, but where participants reported seeing a stimulus, they were strongly biased to report motion in the two most frequently presented motion directions (figure 3). This bias could not be explained as due to any particular ‘response-strategy’. Participants’ perception of real motion stimuli was also influenced by their learned expectations: they showed increased detection performance for the most frequently presented motion directions, and estimated stimuli to be moving in directions that were more similar to the most frequently presented motion directions than they really were (figures 4 - 6). Participants’ estimation behaviour was well described by a model which assumed that they solved the task using a Bayesian strategy, combining a learned prior of the stimulus statistics with their sensory evidence in a probabilistic way (figures 7-9). Further, our model of participants’ behaviour in the presence of a stimulus was able to accurately predict their estimation responses when no stimulus was presented (figure 10).

### Learning the ‘expected’ motion directions

Participants rapidly learned to expect the likely stimuli; within just a few minutes of task-performance. One by-product of such rapid learning was that because participants learned which motion directions were expected within a very few number of trials, it was difficult for us to measure the short term time-course and dynamics of learning (supplementary figure 3). Future work could investigate this using a more complicated distribution of presented stimuli or statistical learning paradigm that produces slower learning of stimulus expectations (Eckstein et al., 2004; Orbán, Fiser, Aslin, & Lengyel, 2008).

Recent studies have shown that rapidly learned expectations influence perception of bistable stimuli (Haijiang et al., 2006; Sterzer et al., 2008). In common with our results, these studies found attractive perceptual biases towards participants’ expectations. However, while these studies looked at perception of relatively complex visual features, such as whether a stimulus was rotating (Sterzer et al., 2008), our experiment looked at perception of simple unambiguous features, which are likely to be processed at a lower level in the visual hierarchy, such as cortical area MT (Newsome, Britten, & Movshon, 1989). Whether similar neural changes are responsible for the effects of expectations on perception of both simple and more complicated stimulus features is an open question.

Our finding, that participants perceived motion in expected directions when nothing was presented, is similar to what has been found in perceptual learning, where after learning participants report seeing dots moving in the trained direction when no stimulus is displayed (Seitz,

Nanez, Holloway, Koyama, & Watanabe, 2005). However, an important difference between our results and what has been reported previously was the time taken for these hallucinations to develop: in the study of Seitz et al. it took around 8 1hr sessions for participants to perceive motion in the trained direction when there was nothing there, while we observed this effect within the first 250 trials. It is interesting to consider whether these visual hallucinations were caused by the same underlying phenomena in both cases. Indeed, elucidating the similarities and differences between the physiological and behavioral effects of different types of learning is an important goal for future research (Seitz & Watanabe, 2005).

### Bayesian model

In our experiment, participants were implicitly asked to learn the statistics of the stimulus directions. In Bayesian terms, this corresponds to learning a prior distribution of the motion stimuli. Bayesian theory (MacKay, 2004) tells us how such knowledge should then be combined with sensory inputs to lead to optimal estimates. Our results can thus be interpreted in the context of 2 questions: 1) are participants able to learn a prior about motion stimuli in the course of our experiment?; 2) is this prior combined optimally with participants’ sensory observations to lead to motion estimates?

We constructed a simple model of participants’ estimation behaviour, which assumed that on each trial they combined their sensory evidence (based on a noisy sensory measurement of motion direction) with a learned prior distribution of ‘expected’ motion directions, in a probabilistically optimal manner (figure 7). For each participant, we chose the width of the likelihood function and shape of the learned prior to maximize the probability of their estimation data being generated by the model. The model provided a good fit of participants’ estimation biases and standard deviations (figure 9). Interestingly, the quality of the fit to the data did not decrease when the width of the likelihood was held constant with presented motion direction (figure 8). On average, the shape of participants’ learned prior (supplementary figure 10) was found to be qualitatively similar to the actual distribution of presented stimuli (figure 2), indicating that they were able to rapidly learn a multi-modal prior distribution of stimulus directions.

In our experiment, the luminance of the two staircased contrast levels (determined by running staircases on the detection performance) were very similar to each other, with a large degree of overlap between them. Therefore, we combined data from both contrast levels for the majority of our analysis. Later, we looked at how participants’ estimation behaviour varied with the stimulus contrast by dividing participants’ estimation responses into ‘low’ and ‘high’ contrast trials, determined by the contrast level of each individual trial, rather than the staircased contrast level that it was a part of (see supplementary materials for details). We found that the average magnitude of participants’ estimation stan-

standard deviations increased for lower contrast levels, along with the magnitude of estimation biases towards the central motion direction (supplementary figure 6).

This is consistent with what we would expect if participants behaved as Bayesian observers. At lower contrast levels, we would expect participants' sensory uncertainty to increase, causing an increase in the standard deviation of estimations. As a result of this, the learned prior would begin to dominate over sensory evidence, causing the magnitude of the estimation biases to increase. While we were not able well fit participants' estimation behaviour at varying contrast using our Bayesian model (as there were too few data points per experimental conditions to well constrain the model) this will be an interesting question for future work.

We reasoned that if our participants were indeed behaving as Bayesian observers, then the prior and likelihood derived from their estimation responses when a stimulus was present should also predict their estimation behaviour when no stimulus was present. This is indeed what we found: the majority of participants' zero-stimulus estimation distributions were well fitted by the model (figure 10). Therefore, while 'hallucinating' motion when none is there will clearly be disadvantageous in most everyday situations (Seitz et al. 2005), in the context of our experiment, it is just what we would expect for an ideal Bayesian observer who sought to minimize their estimation error in the face of perceptual uncertainty.

We compared the Bayesian model with various 'response bias' models, which assumed that participants responded according to different strategies on different trials: either relying entirely on their sensory observations, or on their expectations. These models were worse at describing the estimation data than the Bayesian model (larger *BIC* values; figure 8), leading us to rule them out as an explanation for participants' behaviour in the estimation task.

Our finding that participants responded according to a 'single-strategy' Bayesian model does not necessarily imply that the biases we observed were perceptual in origin. For example, it is possible that participants altered their overall behavioral strategy in order to incorporate knowledge about which motion directions were most likely, while their perception of the stimuli remained unchanged. Indeed, distinguishing between biases that occur at the perceptual or decision-making level is a very difficult task to perform psychophysically (Schneider & Komlos, 2008). However, our modeling work does imply that participants' combined their expectations with their sensory observations in a non-trivial way. Specifically, on each trial participants did not rely solely on either their expectations or their sensory observations, but rather, they made their estimations based on a combination of both of these sources of information. Further, we noted that if the observed estimation biases were due to a change in behavioural strategy, this must have occurred at a largely subconscious level, as most participants were unable to indicate the two motion directions that had been most frequently presented, with a large proportion (9

out of the 12 participants included in our analysis) reporting either that there were equal number of stimuli moving in all directions, or that most of the stimuli were centered around a single motion direction. Also, our personal observations from setting up the experiment is that lab personnel often perceived patterns of moving dots in zero contrast trials, leading us to the conclusion that experimental subjects experienced the same "hallucinations".

## Effect of expectations on performance

We were interested to see whether participants' performance in the detection task was improved for stimuli moving in 'expected' directions. We found that there was a significant increase in participants' detection performance, as well as a significant decrease in reaction time for clicking the mouse during stimulus presentation, for stimuli moving in the most frequently presented motion directions (figure 6 and supplementary figure 7). Although somewhat smaller in magnitude, these effects are similar to what has been reported previously by Sekuler & Ball (1977), who found large improvements in both detection performance and reaction time when participants knew which direction stimuli would be moving in. Such an increase in perceptual sensitivity towards expected stimuli is similar to the effects of selective attention (Doherty et al., 2005; Downing, 1988; Posner et al., 1980), suggesting that the learned expectations led participants to direct selective attention towards the expected stimuli.

## Eye movements

In the experiment of Sekuler & Ball (1977), participants reported that they experienced their eye movements being involuntarily 'pulled' in the direction of the stimulus. It was suggested by the authors that mechanisms controlling eye movements might be capable of responding to very low luminance motion stimuli, and thus, that the resulting eye movements could be used by participants to help them correctly detect stimuli that were otherwise imperceptible.

If this is the case, then it could have also contributed to changes in detection performance and reaction time with motion direction in our experiment. For example, if participants were biased to move their eyes in 'expected' motion directions, then this could result in decreased detection thresholds for these motion directions. However, how such eye movements would influence estimation of motion direction is not so clear. Naively, if participants were biased to move their eyes in expected motion directions, then we might expect this to produce estimation biases away from these directions (as the motion component in this direction would be reduced, relative to the motion of the eye), which is not what we observed. A proper understanding of how extra-retinal eye-movement signals are combined with sensory signals to produce perceptual estimates, is an important area for future work.

## Interaction between tasks

We considered how participants' behaviour in one task could have influenced their behaviour in the other (Jazayeri & Movshon, 2007). Specifically, we asked whether biases in the estimation task could have come about as a result of participants optimizing their behaviour in the detection task. To illustrate how this could happen, consider the case where participants' expectations influenced their detection performance, but not their perception of motion direction. Here, if participants were more likely to detect a stimulus when they perceived it to be moving in 'expected' directions, then this would also cause the estimation distributions to be biased towards these directions when we looked just at trials where a stimulus was detected. However, this bias would disappear when we looked at estimation responses from all trials, regardless of participants' detection responses, which is not what we find experimentally (there was no significant difference between the estimation biases calculated from trials where participants detected stimuli, and from all trials;  $p = 0.71$ , 5-way within-subjects ANOVA).

On the other hand, if, on trials where participants did not detect a stimulus, they treated the estimation task as meaningless and provided random estimation responses, then on average we would still observe a bias towards the expected directions. This could allow participants to respond in a 'self-consistent' way in both tasks (A. Stocker & Simoncelli, 2008): when they have settled on the hypothesis that there is no stimulus present, it makes little sense for them to scrutinize which direction it is moving in. However, as discussed earlier, participants' detection performance varied relatively weakly with motion direction, with a population averaged difference in detection performance of only  $5.9 \pm 1.0\%$  between the two most frequently presented motion directions, and other directions (figure 6). Thus, it seems unlikely that the highly significant variation in estimation biases observed experimentally (varying by  $14.6 \pm 2.9^\circ$  between stimuli moving at  $\pm 16^\circ$  and  $\pm 64^\circ$ ; figure 4a) could be brought about by such small changes in detection performance.

## Expectations and attention

The behavioural effects of sensory expectations have been often linked to those of attention, as both phenomena result in increased perceptual quality for attended or expected stimuli (Doherty et al., 2005; Downing, 1988; Posner et al., 1980; Summerfield & Egner, 2009). In the context of this experiment, it is possible that participants learned to direct feature-based attention towards the most frequently presented motion directions. Therefore, it is worthwhile comparing our results to previous experiments looking at the effects of feature-based attention on motion perception.

Previous studies using transparent motion stimuli, have shown that feature-based attention can modulate how different motion components are perceptually combined, thus altering the perceived directions (Chen, Meng, Matthews,

& Qian, 2005; Tzvetanov, Womelsdorf, Niebergall, & Treue, 2006). For example, Chen et al. (2005) found that attending towards one of two overlapping motion signals reduced the degree of repulsion between the two motion signals, so that the non-attended motion direction was perceived as being closer to the attended motion direction than it would be otherwise. This is consistent with our results, where attending to a particular motion direction resulted in an attractive bias in estimation-responses towards the attended direction. However, in these previous studies, attention acted to select one of two competing motion stimuli, and thus modified the interaction between processing of these different motion signals. Here, we find that when participants 'expect' stimuli to be moving in a particular motion direction, this alters the perceived direction of motion, even in the absence of any competing stimuli.

It is interesting to consider how the perceptual effects that we observed here could be produced by changes at the neural level in the visual cortex. Much modeling work has looked at how visual neurons could encode information about sensory stimuli in the form of probability distributions, both at the single neuron (Deneve, 2008) and population level (Knill & Pouget, 2004; Ma, Beck, Latham, & Pouget, 2006; Pouget, Dayan, & Zemel, 2003). However, at present the evidence for neural encoding of the prior is minimal (Basso & Wurtz, 1997; Platt & Glimcher, 1999; Summerfield & Koechlin, 2008).

On the other hand, recent experiments have shown that expectations of when and where motion stimuli are likely to be presented, can result in increased reliability of neurons in visual area MT (Ghose & Bearl, 2009). In the context of visual attention, numerous studies have shown that selective attention increases the sensitivity of neurons that are tuned towards attended spatial (Spitzer, Desimone, & Moran, 1988; Treue & Maunsell, 1996) or featural (McAdams & Maunsell, 2000; Treue & Martínez Trujillo, 1999) dimensions. Looking specifically at visual motion, electrophysiological studies in macaque MT show that the firing rate of neurons that are tuned towards an attended motion direction are increased relative to neurons that are tuned towards other directions (Treue & Martínez-Trujillo 1999). Therefore, if, in our experiment, participants learned to direct feature-based attention towards expected motion directions, then it is likely that the gain of neurons that were tuned towards these directions was increased. When considered together with our results, this leads to the following questions. First, are the learned priors that seem to be involved in our task encoded directly by gain changes of sensory neurons such as are observed with attention (Dayan & Zemel, 1999; Rao, 2005; Yu & Dayan, 2005a)? Secondly, how are these changes interpreted, or 'decoded', by upstream cortical areas to produce the perceptual biases that we observed (Jazayeri, 2007, 2008; Jazayeri & Movshon, 2006, 2007; Seriès, Stocker, & Simoncelli, 2009)? Finally, an interesting goal for future research is to understand how priors that are learned over a short period of time are incorporated with and used to update long term

priors about the statistical structure of the world (Knill & Pouget, 2004; Weiss et al., 2002).

## Conclusions

We asked whether the statistics of past motion stimuli can modulate perception of new motion directions. This was indeed what we found: participants quickly developed expectations for the most frequently presented directions of motion, and this strongly influenced their perception of simple, unambiguous, visual stimuli, inducing a shift in the perceived direction of stimuli towards expected motion directions as well as hallucinations to see motion when none was presented.

In our work, expectations can be directly interpreted and modelled as Bayesian priors. In a situation like ours where stimuli are presented to only one sensory modality, without conflict or ambiguity, expectations, or Bayesian priors, are often thought to develop slowly over a lifetime of sensory inputs. In contrast, we found that they can be learned rapidly, in a period of a few minutes. Moreover, we showed they are combined with sensory inputs in a way that is compatible with optimal Bayesian inference.

In conclusion, our findings support the idea of a very plastic perceptual system in which prior knowledge is rapidly acquired and constantly used to shape our perceptions towards what we expect to see. Though useful for the system in the face of uncertainty, this plasticity comes at the cost of unconscious illusions and hallucinations.

## Acknowledgments

We would like to thank to Dr M. Oram and Prof A. Thiele for their helpful comments and suggestions. We would also like to acknowledge the helpful and insightful comments provided by two anonymous reviewers on earlier versions of this manuscript. This research was supported by funding from the Engineering and Physical Sciences Research Council and the Medical Research Council of Great Britain.

## References

- Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science*, 218(4573), 697-698.
- Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.*, 5(8), 617-629.
- Basso, M. A., & Wurtz, R. H. (1997). Modulation of neuronal activity by target uncertainty. *Nature*, 389(6646), 66-69.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, 10(4), 433-436.
- Chen, Y., Meng, X., Matthews, N., & Qian, N. (2005). Effects of attention on motion repulsion. *Vision Research*, 45(10), 1329-1339.
- Chun, M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4(5), 170-178.
- Dayan, P., & Zemel, R. (1999). Statistical models and sensory attention. *Artificial Neural Networks*.
- Deneve, S. (2008). Bayesian spiking neurons I: inference. *Neural computation*, 20(1), 91-117.
- Doherty, J. R., Rao, A., Mesulam, M. M., & Nobre, A. C. (2005). Synergistic effect of combined temporal and spatial expectations on visual attention. *J Neurosci*, 25(36), 8259-8266.
- Downing, C. J. (1988). Expectancy and visual-spatial attention: effects on perceptual quality. *Journal of experimental psychology Human perception and performance*, 14(2), 188-202.
- Eckstein, M. P., Abbey, C. K., Pham, B. T., & Shimozaki, S. S. (2004). Perceptual learning through optimization of attentional weighting: human versus optimal Bayesian learner. *J Vis*, 4(12), 1006-1019.
- Faisal, A. A., & Wolpert, D. M. (2009). Near optimal combination of sensory and motor uncertainty in time during a naturalistic perception-action task. *Journal of Neurophysiology*, 101(4), 1901-1912.
- García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research*, 38(12), 1861-1881.
- Ghose, G. M., & Bearl, D. W. (2009). Attention directed by expectations enhances receptive fields in cortical area MT. *Vision Research*, 1-11.
- Haijiang, Q., Saunders, J. A., Stone, R. W., & Backus, B. T. (2006). Demonstration of cue recruitment: change in visual appearance by means of Pavlovian conditioning. *Proc Natl Acad Sci USA*, 103(2), 483-488.
- Jazayeri, M. (2007). Integration of sensory evidence in motion discrimination. *J Vis*, 7(12), 1-7.
- Jazayeri, M. (2008). Probabilistic sensory recoding. *Curr Opin Neurobiol*, 18(4), 431-437.
- Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nat Neurosci*, 9(5), 690-696.
- Jazayeri, M., & Movshon, J. A. (2007). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*, 446(7138), 912-915.
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12), 712-719.

- Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nat. Rev. Neurosci.*, 7(3), 220-231.
- Körding, K., & Wolpert, D. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427(6971), 244-247.
- Lee, T., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*.
- Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat Neurosci*, 9(11), 1432-1438.
- MacKay, D. J. C. (2004). *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press.
- McAdams, C. J., & Maunsell, J. H. (2000). Attention to both space and feature modulates neuronal responses in macaque area V4. *Journal of Neurophysiology*, 83(3), 1751-1755.
- Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, 341(6237), 52-54.
- Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proc Natl Acad Sci USA*, 105(7), 2745-2750.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis*, 10(4), 437-442.
- Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741), 233-238.
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, 109(2), 160-174.
- Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annu. Rev. Neurosci.*, 26, 381-410.
- Rao, R. (2005). Bayesian inference and attentional modulation in the visual cortex. *NeuroReport*.
- Rauber, H. J., & Treue, S. (1998). Reference repulsion when judging the direction of visual motion. *Perception*, 27(4), 393-402.
- Schneider, K. A., & Komlos, M. (2008). Attention biases decisions but does not alter appearance. *J Vis*, 8(15), 31-10.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Seitz, A. R., Nanez, J. E., Holloway, S. R., Koyama, S., & Watanabe, T. (2005). Seeing what is not there shows the costs of perceptual learning. *Proc Natl Acad Sci U S A*, 102(25), 9080-9085.
- Seitz, A. R., & Watanabe, T. (2005). A unified model for perceptual learning. *Trends Cogn Sci*, 9(7), 329-334.
- Seriès, P., Stocker, A. A., & Simoncelli, E. P. (2009). Is the homunculus "aware" of sensory adaptation? *Neural Comput*, 21(12), 3271-3304.
- Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, 240(4850), 338-340.
- Sterzer, P., Frith, C., & Petrovic, P. (2008). Believing is seeing: expectations alter visual awareness. *Curr Biol*, 18(16), R697-698.
- Stocker, A., & Simoncelli, E. (2008). A Bayesian Model of Conditioned Perception. *Advances in Neural Information Processing Systems*, 20, 1409-1416.
- Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nat Neurosci*, 9(4), 578-585.
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9), 403-409.
- Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron*, 59(2), 336-347.
- Treue, S., & Martínez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399(6736), 575-579.
- Treue, S., & Maunsell, J. H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, 382(6591), 539-541.
- Tzvetanov, T., Womelsdorf, T., Niebergall, R., & Treue, S. (2006). Feature-based attention influences contextual interactions during motion repulsion. *Vision Research*, 46(21), 3651-3658.
- van Ee, R. (2005). Dynamics of perceptual bi-stability for stereoscopic slant rivalry and a comparison with grating, house-face, and Necker cube rivalry. *Vision Res*, 45(1), 29-40.
- van Ee, R., van Dam, L. C., & Brouwer, G. J. (2005). Voluntary control and the dynamics of perceptual bi-stability. *Vision Res*, 45(1), 41-55.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat Neurosci*, 5(6), 598-604.
- Yu, A., & Dayan, P. (2005a). Inference, attention, and decision in a Bayesian neural architecture. *Advances in Neural Information Processing Systems*, 17, 1577-1584.
- Yu, A., & Dayan, P. (2005b). Uncertainty, Neuromodulation, and Attention. *Neuron*, 46(4), 681-692.