Vision Research 50 (2010) 460–472

Contents lists available at ScienceDirect

Vision Research

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

Learning motion: Human vs. optimal Bayesian learner

Edgardo J. Trenti [a,*], José F. Barraza [b,c], Miguel P. Eckstein [d]

[a] Departamento de Informática, Facultad de Ciencias Exactas, Universidad Nacional de Salta, Argentina
[b] Departamento de Luminotecnia, Luz y Visión, FACET, UNT, Argentina
[c] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina
[d] Department of Psychology, University of California, Santa Barbara, United States

ARTICLE INFO

Article history:
Received 5 June 2009
Received in revised form 26 October 2009

Keywords:
Learning motion
Perceptual learning
Optimal Bayesian learner
Learning efficiency

ABSTRACT

We used the optimal perceptual learning paradigm (Eckstein, Abbey, Pham, & Shimozaki, 2004) to investigate the dynamics of human rapid learning processes in motion discrimination tasks and compare it to an optimal Bayesian learner. This paradigm consists of blocks of few trials defined by a set of target attributes, and it has been shown its ability to detect learning effects appearing as soon as after the first trial. In the present task a sequence consisting of four patches containing random-dot patterns is presented at four separate locations equidistant from a fixation point. On each trial, the random dots in three patches moved with a mean speed and the fourth, target patch, could move either with *slower* or *faster* mean speed. Observers' task was to indicate what speed, *faster* or *slower*, was present in the display. The mean direction of the target patch was kept invariant along a block of trials. Observers learned the target relevant motion direction through indirect feedback, leading to an improvement in speed identification performance ranging from 15% to 30% which is greater than previously studied contrast defined targets and faces. However, comparison to an ideal learner revealed incomplete or partial learning for the motion task which was lower than previously measured for contrast defined targets and faces. A sub-optimal model that included inefficiencies in the updating of motion direction weights due to memory effects could account for the human learning. Finally, the similarity of the rapid learning effect observed here for motion perception with that found for contrast defined targets for localization and identification tasks could be suggesting a general strategy for learning in the human visual system and some common limitations such as memory.

1. Introduction

Human performance improves with practice for a variety of visual tasks. For example, this effect has been shown for speed and direction discrimination (Saffell & Matthews, 2003), orientation (Matthews, Liu, Geesaman, & Qian, 1999), motion (Ball & Sekuler, 1987; Vaina, Sundareswaran, & Harris, 1995), vernier acuity (Fahle & Edelman, 1993; Mckee & Westheimer, 1978; Saarinen & Levi, 1995), texture segmentation (Karni & Sagi, 1993), spatial displacement (McKee & Westheimer, 1978) and spatial frequency discrimination (Fine & Jacobs, 2000). This improvement of human ability varies widely across tasks (Fine & Jacobs, 2002), and has been associated to a learning process, which may implicate different mechanisms depending on the perceptual task (Goldstone, 1998). In real life, these perceptual tasks involve complex stimuli, which carry a large amount of visual information, but only a portion of this information is relevant to the task. The ecological approach (Gibson, 2000), for example, suggests that perceptual learning consists on identifying those relevant properties from the environment. Several psychophysical studies support this idea and show that perceptual learning improves the human ability to distinguish the relevant information from that is irrelevant and thus, helps to perform a more efficient integration of signal information (Beard & Ahumanda, 1999; Dosher & Lu, 1998; Gold, 2003; Gold, Bennett, & Sekuler, 1999; Hulbert, 2000), perhaps through a reweighting of basic sensory units (Dosher & Lu, 1998). Attention may play a fundamental role in this process (Eckstein, Shimozaki, & Abbey, 2002; Kinchla, Chen, & Evert, 1995; Murray, Sekuler, & Bennett, 2003; Shimozaki, Eckstein, & Abbey, 2003). For example, when a subject first encounters a complex visual task, there is high uncertainty about what are the relevant cues that will allow the subject to perform the task successfully. However, during practice, the subject learns those informative cues and automatically gives more importance to these cues by attending them and ignoring those that are irrelevant. This mechanism of perceptual learning has been referred to as learning through reduction in uncertainty or learning through attention optimization (for a review about learning mechanisms, see Goldstone, 1998), and is supported by a number of studies suggesting that, in this process, attention allows the

* Corresponding author.
E-mail address: jtrenti@unsa.edu.ar (E.J. Trenti).

0042-6989/$ - see front matter
doi:10.1016/j.visres.2009.10.018

system to differentially weight and/or select sensors coding task relevant information (Eckstein et al., 2002; Jacobs, 2009; Kinchla et al., 1995; Murray et al., 2003; Shimozaki et al., 2003).

It has been shown recently, that this process of perceptual learning may occur in few trials (Abbey, Pham, Shimozaki, & Eckstein, 2008; Eckstein, Abbey, Pham, & Shimozaki, 2004; Peterson, Abbey, & Eckstein, 2009). The authors developed a new experimental paradigm (optimal perceptual learning, OPL) to systematically study rapid perceptual learning processes by allowing the comparison to that of an optimal Bayesian algorithm (see also, Jacobs, 2009). They studied the improvement of humans' ability in the localization of elongated targets with different orientations and polarities. Results showed that humans can significantly improve their localization performance as they reduce the uncertainty about the orientation and contrast polarity of the target. They showed that this process can occur in just four trials but that this rapid human learning is slower and incomplete with respect to the Bayesian learner (but see Peterson et al. (2009) for learning comparable to that of a Bayesian learner).

In this paper, we propose to use this paradigm to study whether this learning process applies to motion perception. In particular, we are interested on testing whether an observer's learning of the task's relevant motion direction improves speed discrimination judgments. Motion is a very important component of vision. Not only because animals move through the environment but because a number of salient objects move in our visual field. Estimating correctly the parameters of visual motion may play a critical role on the success in performing relevant tasks such as segmenting the image, avoiding dangerous objects, or the identification of relevant features in a complex scene, among others. Locally, those motion parameters are speed and direction; however, it is well known that combining several different local motion signals can also produce a percept of coherent motion in a single direction (Williams & Sekuler, 1984). In these displays, one can also perceive a global speed that depends on this resultant direction, which may be predicted by a vector averaging process (Curran & Braddick, 2000). Because, these two parameters are always present in a moving object, it is of interest to understand how they interact, and whether the rapid reduction of uncertainty about one parameter (direction of motion) can affect ability to make discrimination judgments along the other dimension (speed). As an ecologically relevant example one might ask whether learning the likely motion direction of a ball flying in a crowd of soccer players might improve the observers' estimate about the speed of the ball.

In order to investigate whether learning motion direction improves speed discrimination we devised an experiment that also allows us to explore the algorithms that underlie this process. The experiment was designed so that an optimal Bayesian learner can be defined. This will allow us to compare the performance of the human subjects to that of an ideal observer that is established as a standard (Eckstein, Drescher, & Shimozaki, 2006; Eckstein et al., 2004; Tjan & Legge, 1998; for a review about this topic, see Kersten & Yuille, 2003, Geisler, 2003), and provides a framework that considers the complexity of the task and the available stimulus information (Liu, Kersten, & Knill, 1999; Liu, Knill, & Kersten, 1995).

2. Theory

2.1. OPL: optimal perceptual learning paradigm

We used the OPL paradigm (Eckstein et al., 2004) to explore rapid changes of performance of motion discrimination due to attention optimization. In the present OPL task an image sequence consisting of four patches containing random-dot patterns is presented at four separate locations equidistant from a fixation point. On each trial, the random dots in three patches moved with a mean speed and the fourth, target patch could move either with *slower* or *faster* mean speed (with 50% probability faster and 50% probability slower speed). The speed of each dot was independently sampled from a Gaussian distribution with the patch's assigned mean (medium speed or faster/slower speed). The observers' task was to determine whether the patch with fast or slow mean speed dots was present in the display. Trials were blocked into groups of four, which we will refer to as learning blocks. For each learning block, a mean direction of motion for the dots in the target patch was selected from the four possible cardinal directions (with equal probability) and kept constant for the target patch throughout the learning block. The motion direction for each dot of a patch was independently sampled from a Gaussian distribution with a mean corresponding to one of the four cardinal directions assigned to the patch. The spatial position of the patches (target and non-target) varied randomly from trial to trial. After the observer's decision on each trial, feedback was given to the observer about the spatial location containing the target patch of dots but no feedback was given about the direction of motion of the target patch. At the end of the learning block of trials subjects were asked to indicate the direction of motion of the dots in the target patch for that block of trials. After such decision, feedback was given indicating whether their decision was right or wrong. Fig. 1 represents schematically the timeline of the procedure. Human performance discriminating the speed of the target is quantified by calculating the proportion correct answers for each learning trial (1st–4th).

2.2. Why would performance in the speed judgment improve with learning trial?

On the first trial of a learning block, observers did not have any information of the mean direction of dot motion of the target patch and thus have to either base their speed decision on the sensory data from all of the patches (task relevant target patch and task irrelevant non-target patches) or choose a subset of the patches. After the first learning trial feedback is presented about the location of the target patch which provides the observer with information of the likely direction of motion of the target patch for that learning block of trials. This information could potentially be used in the next trial to base their speed judgment decision on the task relevant motion patch and exclude the task irrelevant motion patches that bring unnecessary noise into the decision. Performance in the speed judgment could potentially improve with learning trials due to the more optimal integration of sensory information across motion patches. Identification of the target patch among the four patches was non-trivial for humans due to the addition of motion direction noise.

2.3. Optimal Bayesian learner

The optimal perceptual learning paradigm allows us to postulate the optimal Bayesian learner for the task. The algorithm calculates the posterior probability for each hypothesis, in this case, slower or faster, by using all the available data on the stimulus and chooses the speed with the higher posterior probability. The posterior probability of the gth hypothesis ($g = s$ *slower*, $g = f$ *faster*) can be related through the Bayes' rule to the likelihood of the data given the presence of the gth speed:

$$P(S_g|data) = P(S_g)P(data|S_g)/P(data) \quad (1)$$

where $P(S_g|data)$ is the posterior probability of the signal being present with the gth speed given the data, $P(data|S_g)$ is the likelihood of the data given target presence with the gth speed, and

$P(S_g)$ is the prior probability of the signal being present with the gth speed, which is 0.25 for the current experiment. $P(data)$ is the probability of the data, which is independent of speeds and can be replaced by 1 without affecting the outcome of the decisions. Therefore, the computation of the posterior probability is reduced to the computation of the likelihood, $P(data|S_g)$ (see Fig. 2).

Because each patch contains multiple dots and frames (temporal samples), whose speed and motion direction are sampled independently from a Gaussian distribution we can describe the set of speed measurements for all dots and frames in a patch with the vector **s** and its N motion directions by a vector **d**. The joint likelihood of observing a patch j of N dots for H frames moving at speed

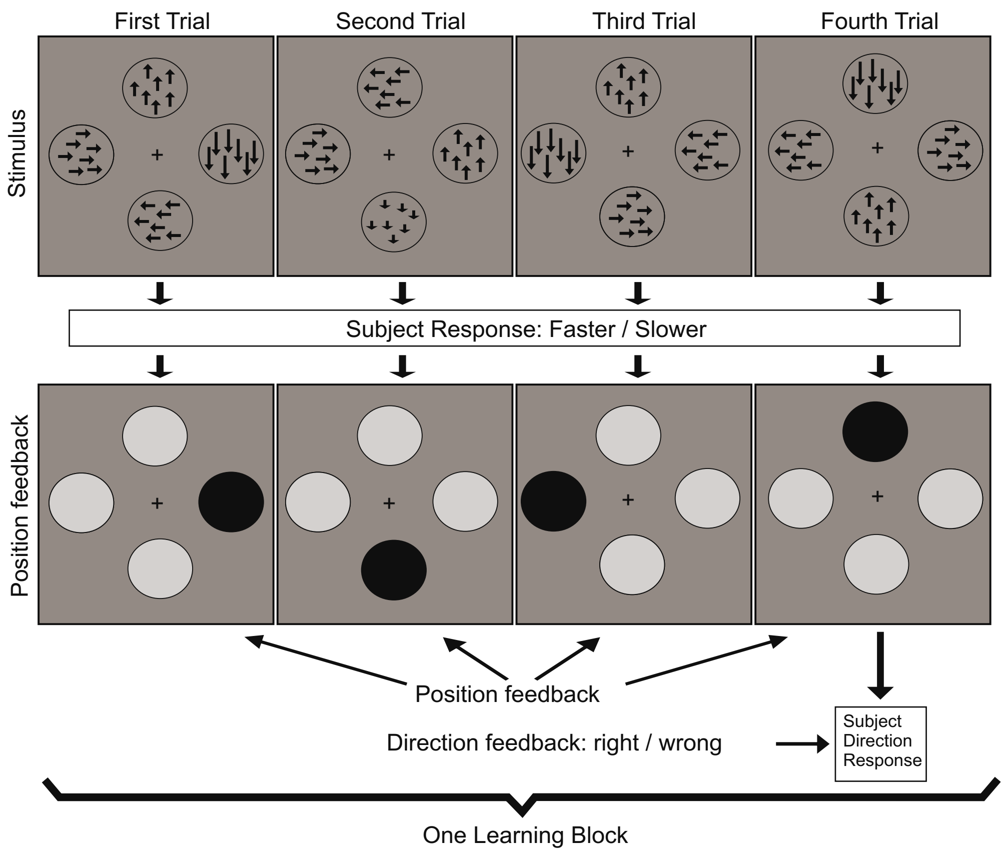

Fig. 1. Structure of a learning block in our OPL paradigm. The arrows denote speed and direction of dots. Note that in the real stimuli these signals are noisy. In this example, the target direction was π. On each trial, after the stimulus presentation, the subject has to respond about target speed, which can be faster or slower. Then, the subject receives a positional feedback. The rest of the trials have the same structure, except that after the fourth trial, the subject is asked to report the direction of the target and then, feedback about subject response is provided to the observer.

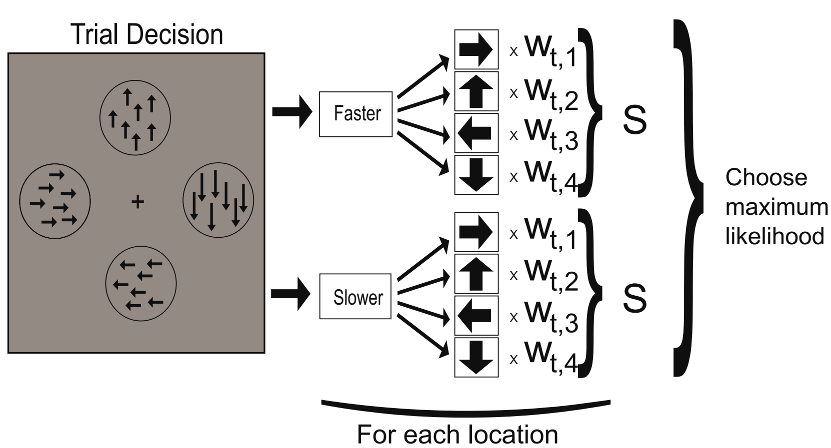

Fig. 2. Scheme of the ideal observer decision for choosing the speed in each trial. To decide if the signal patch is faster or slower the algorithm summates the conditional densities of data for each weighted direction and location. The Algorithm calculates a likelihood for faster and a likelihood for slower and chooses the maximum likelihood to make the decision.

s, and in direction **d** given a mean speed S_g and a mean direction of motion D_i is given by:

$$P(\mathbf{s}_j, \mathbf{d}_j | S_g, D_i) = \prod_{n=1}^{N} \prod_{h=1}^{H} [P(\mathbf{s}_{j,n,h} | S_g) P(\mathbf{d}_{j,n,h} | D_i)] \quad (2)$$

where the probabilities are multiplied across N statistically independent dots and H frames and where $P(\mathbf{s}_{j,n,h}, \mathbf{d}_{j,n,h} | S_g, D_i) = P(\mathbf{s}_{j,n,h} | S_g) P(\mathbf{d}_{j,n,h} | D_i)$ given that motion speed and direction are statistically independent. Eq. (2) expresses the likelihood of the dots at one patch. The probability of the data at the four locations (four patches) given that the patch at the jth location has a faster speed and upward direction and the remaining three patches have the other three remaining directions and a medium speed can be calculated as:

$$\begin{aligned} &P(\mathbf{s}_1, \mathbf{d}_1; \mathbf{s}_2, \mathbf{d}_2; \mathbf{s}_3, \mathbf{d}_3; \mathbf{s}_4, \mathbf{d}_4 | S_f, D_1; S_m, D_2; S_m, D_3; S_m, D_4) \\ &\quad = [P(\mathbf{s}_1 | S_f) P(\mathbf{d}_1 | D_1)][P(\mathbf{s}_2 | S_m) P(\mathbf{d}_2 | D_2)][P(\mathbf{s}_3 | S_m) P(\mathbf{d}_3 | D_3)] \\ &\qquad \times [P(\mathbf{s}_4 | S_m) P(\mathbf{d}_4 | D_4)] \end{aligned} \quad (3)$$

Here, $P(\mathbf{s}_1|S_f)$ is the probability of observing a speed at the patch at location 1 ($\mathbf{s}_1$) given that patch is sampled from the fast speed (S_f) and $P(\mathbf{d}_1|D_1)$ is the probability of observing the dots moving in a direction **d**, at the patch at location 1, ($\mathbf{d}_1$), given that it is sampled from a distribution with a mean direction D_1. The other terms are similarly defined.

Eq. (3) considers only one possible scenario of the random assignments between speed, directions and patch locations (note that we have only specified the subscript j corresponding to the patch and ignored the subscripts n and h for simplicity). Calculating the likelihood of the data given that a speed is faster or slower, the Bayesian ideal observer needs to consider the uncertainty in the OPL paradigm about the location (varying from trial to trial within a learning block) and motion direction (varying across learning blocks) of the relevant dots. Thus, the likelihood of the data given that a patch of dots is moving at a faster or slower speed needs to be calculated by summing across likelihoods of the mutually exclusive events (possible location of the dot patch containing the fast or slow speed and assignment of the different motion directions of the dots of a patch to each location):

$$P(data_t | S_f) = \sum_{i=1}^{dir} w_{i,t} \sum_{j=1}^{loc} P(\mathbf{s}_{j,t} | S_{f,t}) P(\mathbf{d}_{j,t} | D_{i,t}) A_{k \neq i, h \neq j} \quad (4)$$

where the summations are over locations ($loc = 4$) and directions ($dir = 4$). The variable $w_{i,t}$ is the weight in the tth trial for the ith motion direction and is central to the learning ideal learner. $P(\mathbf{s}_{j,t}|S_f)$ is the probability to observe in the patch at the jth *location* the speed s given that the speed is the *faster*. $P(\mathbf{d}_{j,t}|D_i)$ is the joint probability to observe in the patch at the jth *location* the direction d given that the direction is D_i, and $A_{k \neq i, h \neq j}$ is defined as

$$A_{k \neq i, h \neq j} = \prod_h P(\mathbf{s}_h | S_m) \sum_{l=1}^{(dir-1)!} \left[\prod_{h,k} P\left(\mathbf{d}_h \middle| D_{\sigma_{l,k}}\right) \right] \quad (5)$$

A is the product of the probabilities to observe in the patch h a speed **s** given the distribution of medium speeds S_m and a direction d given the distribution of direction D_k, by considering all the possible permutations ($\sigma_{l,k}$) over the patches locations and directions different from those of the target, which in this case are $(dir - 1)!$ permutations. To illustrate this calculation, we developed Eq. (3) for the case in which $i = 1$ and $j = 1$, which corresponds to a target whose direction is 0° (North), located on the upper patch. The expression for this case is

$$\begin{aligned} A_{k \neq 1, h \neq 1} = &P(s_2|S_m)P(s_3|S_m)P(s_4|S_m)[P(\mathbf{d}_2|D_W)P(\mathbf{d}_3|D_S)P(\mathbf{d}_4|D_E) \\ &+ P(\mathbf{d}_2|D_W)P(\mathbf{d}_3|D_E)P(\mathbf{d}_4|D_S) + P(\mathbf{d}_2|D_E)P(\mathbf{d}_3|D_W)P(\mathbf{d}_4|D_S) \\ &+ P(\mathbf{d}_2|D_E)P(\mathbf{d}_3|D_S)P(\mathbf{d}_4|D_W) + P(\mathbf{d}_2|D_S)P(\mathbf{d}_3|D_W)P(\mathbf{d}_4|D_E) \\ &+ P(\mathbf{d}_2|D_S)P(\mathbf{d}_3|D_E)P(\mathbf{d}_4|D_W)] \end{aligned} \quad (6)$$

where the subscript W, S, E denote the directions west, south, and east.

These six terms represent all the permitted permutations over the directions different from *North*, for each patch different from Patch 1. Note from Eq. (4), that the complete expression contains sixteen of these terms for each dot in the patch, which implies an extremely costly computation. Moreover, several numerical complications appear in the algorithm due, for example, to the range of directions used in the experiment (problem of circularity), or to the way we compute dot's velocity, which may produce correspondence noise. These issues are discussed in detail in Appendix.

How does the ideal observer learn? The algorithm begins with the uncertainty about the motion direction and patch containing the slow or fast speed. This uncertainty is denoted by equating the priors (weights) affecting each direction in Eq. (5). Thus, on the first trial, the ideal Bayesian learner makes a decision about the speed by integrating information equally across all motion directions and patch locations. However, since the motion direction of the signal patch is fixed for a block of trials, in subsequent trials, the feedback information is used to modify the priors' weights such that the response of the ideal observer is biased by the direction carrying the signal. To do this, the ideal observer takes the dots of the patch indicated by the feedback to contain the signal and estimates the probability that their directions belong to each of the four possible direction distributions such is expressed with the following formula:

$$w_{i,t} = \prod_{t'=1}^{t-1} P(\mathbf{d}_{l',t'} | D_i) \quad (7)$$

where l' is the position given by feedback, and t is the current trial. Therefore, the weights modify their values depending on the correspondence between data obtained on each trial in the location signaled by the feedback with the direction distribution i. These values are then normalized and multiplied by the likelihood of each location to decide whether the patch carrying the signal is *faster* or *slower*. The Fig. 3 shows an example, with an exaggerated level of noise, of the evolution of weights for the particular direction 0.

3. Simulations

The performance of the model was calculated by using Monte Carlo simulations. The purpose was to analyze the behavior of the ideal observer, in order to obtain some characteristics regarding the form in which a human observer should behave. The simulations were performed for directional noise ranging between $\pi/5$ and 2π, with steps of $\pi/5$, and speed noise ranging between 1 and 6 deg/s with increments of 0.5 deg/s. The number of trials was 4000 per point, because of the computational cost of the simulation. Each point in the surface of Fig. 4 took approximately 1 h and 15 min running on a Pentium Dual Core with 512 Mb of RAM. Fig. 4 shows the proportion correct for speed discrimination as a function of speed and directional noise, for the four learning trials (rows from top to bottom), and for two displays with different numbers of dots per patch (columns). The shape of the response surface obtained for 10 and 20 dots are similar except that higher proportion correct responses were obtained for 20 dots, which is expected since it contains more information. Both columns show learning in speed discrimination along trials. This effect is stronger for low levels of directional noise, and for the

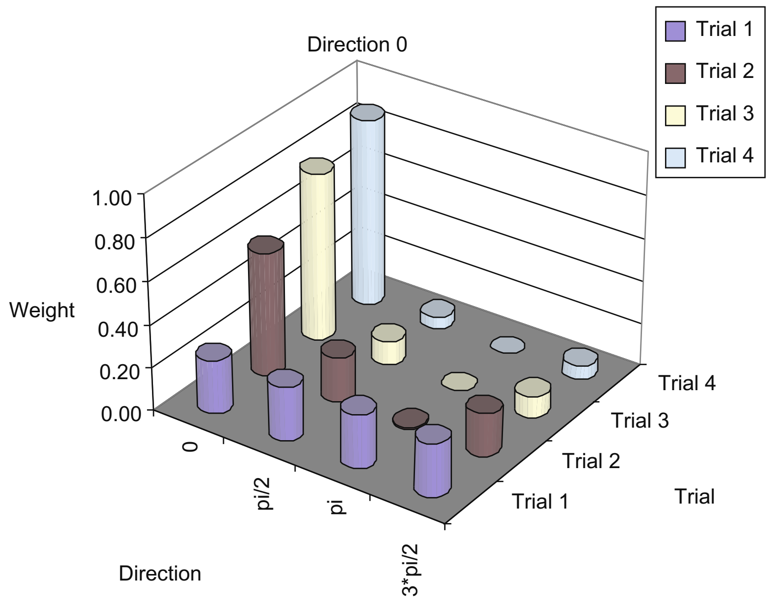

Fig. 3. Weight evolution for the four cardinal directions as a function of learning trial. Results correspond to the case in which the target moved with a mean direction of 0.

ideal Bayesian observer it occurs almost exclusively between first and second learning trials. Plots show that in the first learning trial, the proportion correct does not depend on directional noise because there is maximum uncertainty about direction. However, in the second trial, after the reduction in uncertainty, the relevant motion direction is directly affected by the amount of motion direction noise. Low amplitudes of motion direction noise lead to more certainty in the second learning trial about the relevant motion direction.

Interestingly, the proportion of correct responses for direction was close to unity even for high levels of noise in speed and direction. This high performance is related to the means of the distributions of direction differing by $\pi/2$, which makes them easily discriminable for the Bayesian ideal observer.

3.1. Efficiency

A comparison between human and ideal performance can be made through the measurement of efficiency. The efficiency is the squared ratio between the signal contrast that ideal observer requires to obtain the same level of performance as the human observer, and the human signal contrast (Barlow, 1980). We ran simulations varying the speed contrast, which was defined as $c = \Delta s/s$, where Δs is the increment/decrement for the differential speed, and s is the average speed, for one level of directional noise per subject (0.8717 for MD, 0.8260 for JT, and 0.7002 for PB) and a speed noise of 0.5 deg/s for the three observers. Therefore, the efficiency is expressed as $\eta = (c_{\mathrm{ideal}}/c_{\mathrm{human}})^2$, and this is applied to each trial. The denominator is the speed contrast selected for the human psychophysical experiment. The numerator is found by varying the speed contrast in the simulations for the ideal observer to achieve the same value of proportion correct reached by the human observer for that learning trial. Each point was calculated from 4000 trials of the ideal observer. Proportion correct data were fitted with a logistic function.

3.2. Learning efficiency

Although the absolute efficiency allows us to quantify the performance of a human observer respect to an optimal algorithm, it does not tell much about the effects of learning. Abbey and collaborators (2008) introduced a quantity they term the learning efficiency, which is independent from the absolute efficiency that measures changes in threshold relative to the ideal observer and a non-learning observer that is otherwise ideal. The learning efficiency is defined as the ratio between two differences in threshold energy: the elevation of threshold representing the contribution of learning to that observer's performance, and the elevation of threshold that represents the contribution of prior images to the ideal observer (for details see Abbey et al., 2008)

$$LE_i = 100 \times \frac{\Delta E_{Obs}(i)}{\Delta E_{IO}(i)} \tag{8}$$

Therefore, the learning efficiency measures the observer's threshold elevation respect to the ideal observer's threshold elevation.

4. Psychophysical experiments

4.1. Methods

Subjects had to learn the direction of the patch containing the target, through the four trials of a learning block. The stimulus consisted of four patches of random dots, each one of which moved to one of the four cardinal directions (0, $\pi/2$, π and $3\pi/2$ radians). Each dot subtended a visual angle of 10″. Matlab and Phsychtoolbox (Brainard, 1997; Pelli, 1997) were used to develop and display the stimuli. The patches were circular with a diameter of 5°, and were located 10° away from the center of the screen, over abscissa and ordinate axes. Stimuli were displayed on a 19′ flat CRT monitor at a viewing distance of 67 cm. The frame rate was 60 Hz and the stimuli duration was 12 frames. The dots' lifetime consisted of all 12 frames. The background luminance of the stimulus was 47 cd/m^2. Each dot moved with a speed, which was chosen from Gaussian distributions with means 3.2, 4, and 4.8 deg/s for *slower* and *faster* speed respectively. The standard deviation was 0.5 deg/s in all cases. The direction of each dot was sampled from a Gaussian distribution which means corresponded to the four cardinal directions and their standard deviations were obtained for each subject from a preliminary experiment. In this experiment we measured the amount of noise necessary for the subject to

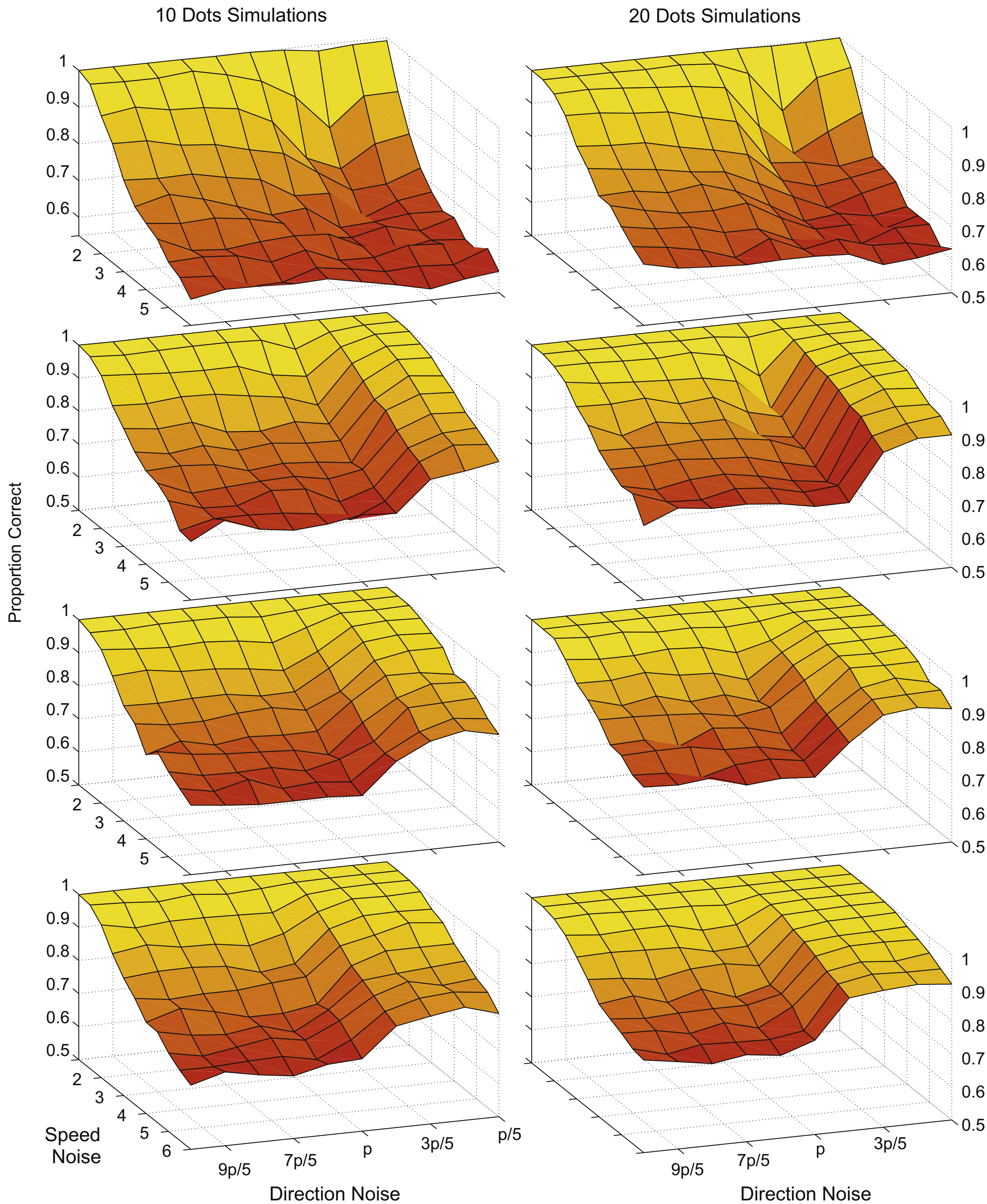

Fig. 4. Proportion correct as a function of speed and directional noise, for both number of dots. Different panels in each column represent trials 1–4.

get a performance (proportion correct) of 0.75 in a direction discrimination task. Subjects had to discriminate between two motions whose directions differed in $\pi/2$. The values in radians for each subjects are: 0.8717 for MD, 0.8260 for JT, and 0.7002 for PB.

At the beginning of a learning block, a message was displayed at the center of the screen asking for the subject to click the mouse to initiate the sequence of trials. On each trial, the location of each direction was chosen randomly and independently from the rest of the trials. The target speed could also be *faster* or *slower* independently from the speed of the other trials in the block. Importantly, the speeds *faster* or *slower* always accompanied the same target direction in a given block of trials.

At the end of each trial the subject was asked to report whether the target speed was *faster* or *slower* by clicking on the corresponding button, which were displayed on both sides of the fixation point. Then, a positional feedback was given to the observer. At the end of each block, four buttons with the cardinal directions were displayed in random order along the width of the screen, and the subject had to indicate by clicking on one of them the target direction. Feedback to the subject was given at the end of the block of learning trials to indicate whether the answer was right or wrong.

Three observers participated in the experiments, two naïve observers, and one of the authors. All of them had normal or corrected to normal acuity. Viewing was binocular with normal pupil. Subjects were instructed to perform the speed discrimination task while fixating on a marker located on the center of the screen.

The experiment consisted on one thousand blocks of learning trials. Thus, each subject performed 4000 speed-discrimination-trials (see Fig. 5).

4.2. Psychophysical results

The experiment was performed for two display conditions consisting of different total number of dots per patch: 10 and 20 dots. Fig. 6 shows proportion correct as a function of learning trial for the three observers. Average proportion correct in the speed discrimination improved along the four trials of the learning blocks for all three observers. The improvements in proportion correct from first to fourth learning trials were (for subjects MD, JT, and PB respectively) 20.24%, 14.66%, and 30.70 %, for the 20 dots per patch display and 20.03%, 21.60%, and 26.92% for the 10 dots per patch. Confidence intervals and t-tests were calculated to compare the proportion correct obtained along learning trials and to compare data obtained with both display conditions. We applied a Bonferroni correction with $n = 10$ to the significance level used in the multiple comparisons. The three observers showed significant differences between first and second trials for both conditions (p-value < 0.0001). Data of Observer MD showed no significant differences among subsequent trials. For observer PB, there was a significant difference between second and third trials of the 20 dots condition (p-value < 0.0001). Finally, for observer JT, a significant difference was found between third and fourth trials of the 10 dots condition (p-value = 0.0031). On the other hand, for observer MD the data of proportion correct between experiments with 20 and 10 dots, showed no significant differences along the four trials. For observers JT and MD the comparison between the results obtained with 20 and 10 dots showed significant differences only for the third trial (p-value = 0.0037 and p-value = 0.0038, respectively).

4.3. Comparisons to the ideal Bayesian observer

Fig. 6 shows a comparison between the human observer and the Bayesian ideal observer. For this comparison the speed contrast in the simulation for the ideal observer was adjusted to match the performance of each observer in the first learning trial. The rapid growth of speed discrimination performance for the ideal observer in comparison with the human observer could be explained by the rapid growth of the weight corresponding to the target motion direction due to the fact that the ideal observer has no uncertainty about the data presented in the patch signaled by the feedback (perfect memory). In other words, the ideal observer is more efficient at integrating motion direction information than the human observers. For example, a directional noise that leads human observers to approximately 75% of performance in motion direction discrimination, allows the ideal observer to perfectly identify (100%) the relevant motion direction after the 1st learning trial. This rapid reduction in uncertainty about motion direction for the ideal observer results in rapid and abrupt performance improvement in the speed discrimination task after the 1st learning trial.

Fig. 7 shows the efficiency for each observer and number of dots. The plots show that the efficiency is very low for all cases, as was expected from the analysis of the performance of the ideal observer. The shape of the curves shows an incomplete or partial learning signature (Abbey et al., 2008; Eckstein et al., 2004) for the three observers and a pronounced decrement of efficiency from

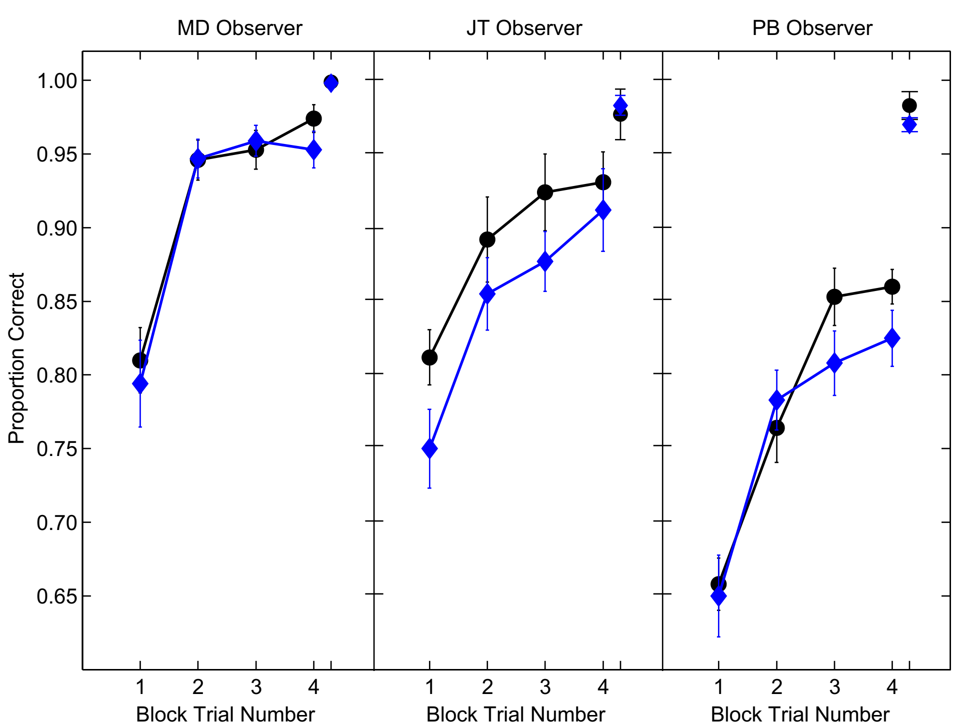

Fig. 5. Proportion correct as a function of trial number for the speed discrimination task. The blue and back curves represent data obtained with 10 and 20 dots respectively. Each panels show the results of one subject.

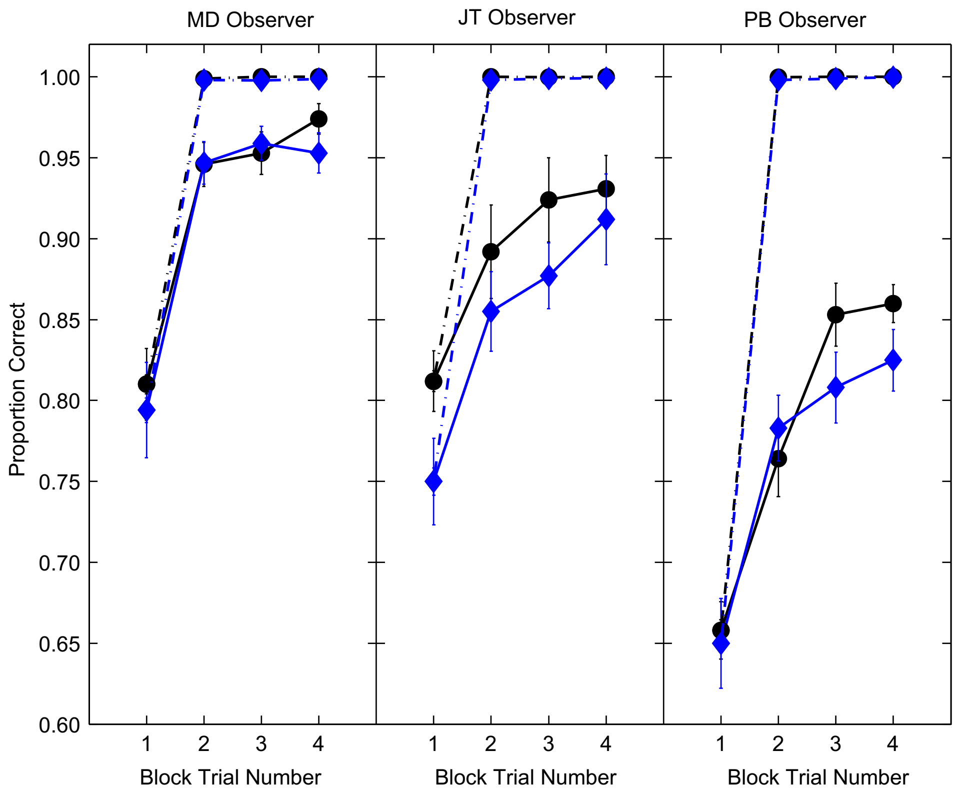

Fig. 6. Comparison between ideal and human observers' performance. The performance of ideal and human observers are leveled at the first trial. The ideal observer uses this speed contrast for the rest of the learning block. The blue and black curves represent 10 and 20 dots per patch.

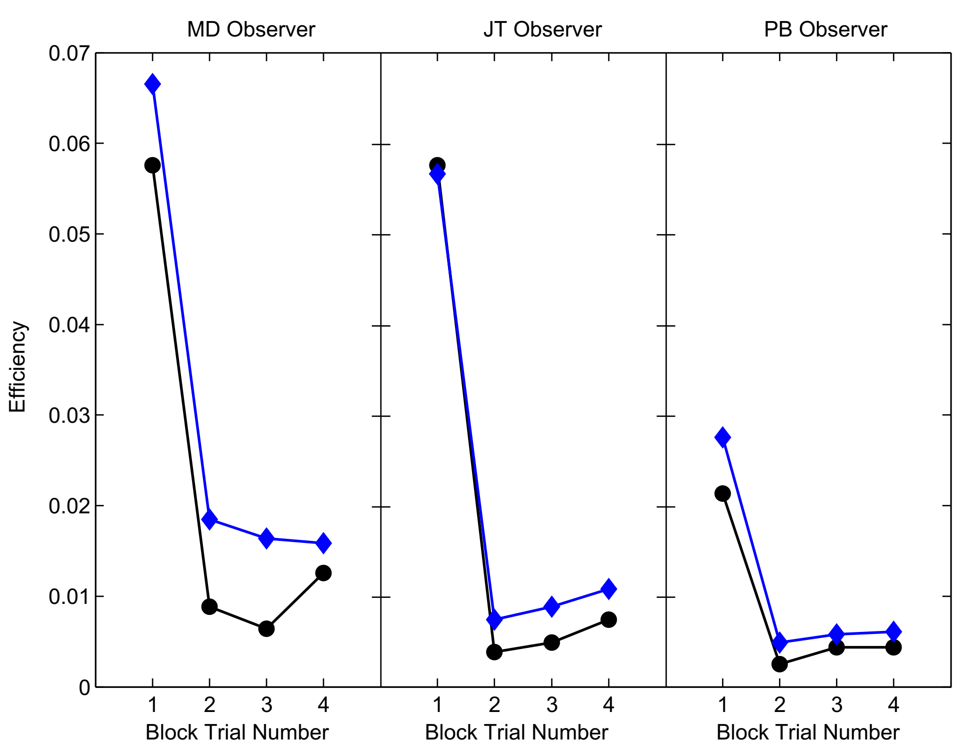

Fig. 7. Absolute efficiency as a function of trial number for the three observers. The blue and black curves represent the efficiency for 10 and 20 dots respectively.

the first to the fourth trial of 78.22% and 76.15% for 20 and 10 dots respectively for the observer MD; 87.16% and 80.91% for the observer JT; and 79.56% and 77.92% for the observer PB.

Fig. 8 shows, for both numbers of dots, the average learning efficiency as a function of trial number. Note that the abscissa begins with the second trial. This is because the learning efficiency is

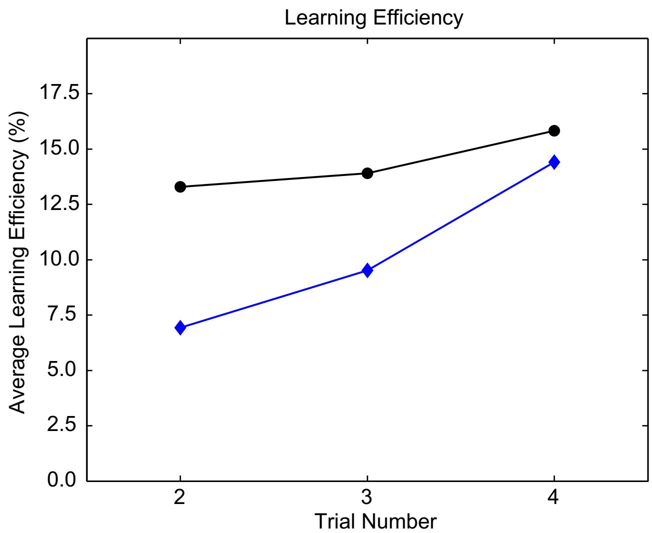

Fig. 8. Learning efficiency as a function of trial number for the three observers. The blue and black curves represent the efficiency for 10 and 20 dots respectively.

indeterminate in Trial 1, which reflects the fact that learning cannot occur in the first trial. Results show that the learning efficiency increases with trial number consistently with previous results (Abbey et al., 2008). Interestingly, data show that learning efficiency is higher for 10 than for 20 dots, which suggests that increasing the number of dots would not make the speed discrimination task easier, such is the case of contrast in a target localization task (Abbey et al., 2008).

4.4. Sub-optimal observer algorithms

Due to the inefficiency of human learning relative to the ideal learner we pursued a number of sub-optimal models that degrade the ideal Bayesian observer. The goal is to determine whether with the sub-optimalities the models can capture human performance.

4.4.1. Partial memory for location containing the signal

One possible method to degrade the ideal observer is to introduce uncertainty in the position signaled by the feedback in the priors update. This model would assume that human observers cannot precisely remember which location contained the signal as indicated by the feedback and updated the trial to trial weights (priors) for each direction based on a weighted sum of sensory evidence across all four locations rather than the location indicated by the feedback.

We can modify Eq. (7) to introduce this uncertainty such that the weight for a direction i is determined by a weighted sum of the sensory evidence for that direction (probabilities of the data at each of the four patches given the distribution of directions D_i) across all locations:

$$w_{i,t} = \prod_{t'=1}^{t-1}[fmP(\mathbf{d}_{l',t'}|D_i) + fP(\mathbf{d}_{1',t'}|D_i) + fP(\mathbf{d}_{2',t'}|D_i) + fP(\mathbf{d}_{3',t'}|D_i)] \quad (9)$$

where l' is the patch position given by the feedback and 1′, 2′, and 3′ are the rest of the patches. fm and f are factors representing the amount of memory such that $fm + 3f = 1$. Therefore, if $fm = 1$ the Eq. (9) is reduced to Eq. (7) and means no uncertainty or perfect memory. On the other side, $fm = f = 0.25$ represents null memory and the model will not learn because any direction is favored by the feedback along trials. Between these two situations, a fm larger than 0.25 and smaller than 1 may represent a real memory. We performed a series of simulations and fit the parameters of the model to the human data. Parameter settings $fm = 0.4$ and $f = 0.2$ resulted in similar amount of learning for the model and the human observers, however, the shape of the learning curve is drastically different from that of humans. The main characteristic of human performance is the rapid increase between 1st and 2nd trials and a moderate increase in the subsequent trials (see Fig. 9) which is not captured by the present model. Note that the present model suggests that observers may fail in remember the location signaled by the feedback. This seems to be unlikely since the feedback is a strong and noise free signal that observers easily remember (this assumption is supported by the observers' reports and our own observation). Therefore, the fact that the model cannot account for the data may be suggesting that the low efficiency found in human observers would not be determined by this source of sub-optimality.

4.4.2. Cautious updating with noisy memory model

We considered two more sources of sub-optimality. First, in the model we name cautious updating model (CM), the observer updates the priors based on the information at the chosen patch but only on those trials in which the feedback signals that patch. In other words, the observer only updates its priors in correct trial. This sub-optimal model has been used previously to capture human learning for contrast defined targets s (Abbey et al., 2008; Eckstein et al., 2004). The weights are calculated in each trial as

$$w_{i,t} = \begin{cases} \prod_{t'}^{t-1} P(\mathbf{d}_{l',t'}|D_i) & \text{for correct speed trials} \\ w_{i,t-1} & \text{for incorrect speed trials} \end{cases} \quad (10)$$

The other source of sub-optimality we investigated is to assume that finite memory would increase the internal noise in the representations of sensory evidence of motion direction. Because the location feedback follows the extinction of the dot display, any calculation of the likelihoods of the data at the feedback location given a motion direction must rely on a memory representation of the data. We model this finite memory by adding a component of internal noise (IN) to the data used to update the weights for the different motion directions, such is shown in Eq. (11).

$$w_{i,t} = \prod_{t'}^{t-1} P(\mathbf{d}_{l',t'} + \mathbf{q}|D_i) \quad (11)$$

where $\mathbf{q}$ is a vector containing the simulated internal noise, whose distribution was normal, with $\mu = 0$ and σ the variable of the simulations. Therefore, combining these two equations we can calculate the weights for each trial in the case that both sources of sub-optimality are present

$$w_{i,t} = \begin{cases} \prod_{t'}^{t-1} P(\mathbf{d}_{l',t'} + \mathbf{q}|D_i) & \text{for correct speed trials} \\ w_{i,t-1} & \text{for incorrect speed trials} \end{cases} \quad (12)$$

Fig. 10 shows, for the three observers, the best fit of the two models (CM in red and IN in green)[1] and their combination (yellow) to the human data (blue). The plots show that neither CM nor IN models can account well for the limited learning of humans, except for the case of Observer MD, whose performance would seem to be well-fit by the noisy memory model. For observers JT and PB, the model that includes both updating only on correct trials and noisy memory (CM + IN) fits the human data best (Fig. 10). The best fits were obtained for values of noise of: $[59/128]\pi$ for MD, $[61/128]\pi$ for JT, and $[66/128]\pi$ for PB. Importantly, the

[1] For interpretation of color in Figs. 5–8, 10 and 11, the reader is referred to the web version of this article.

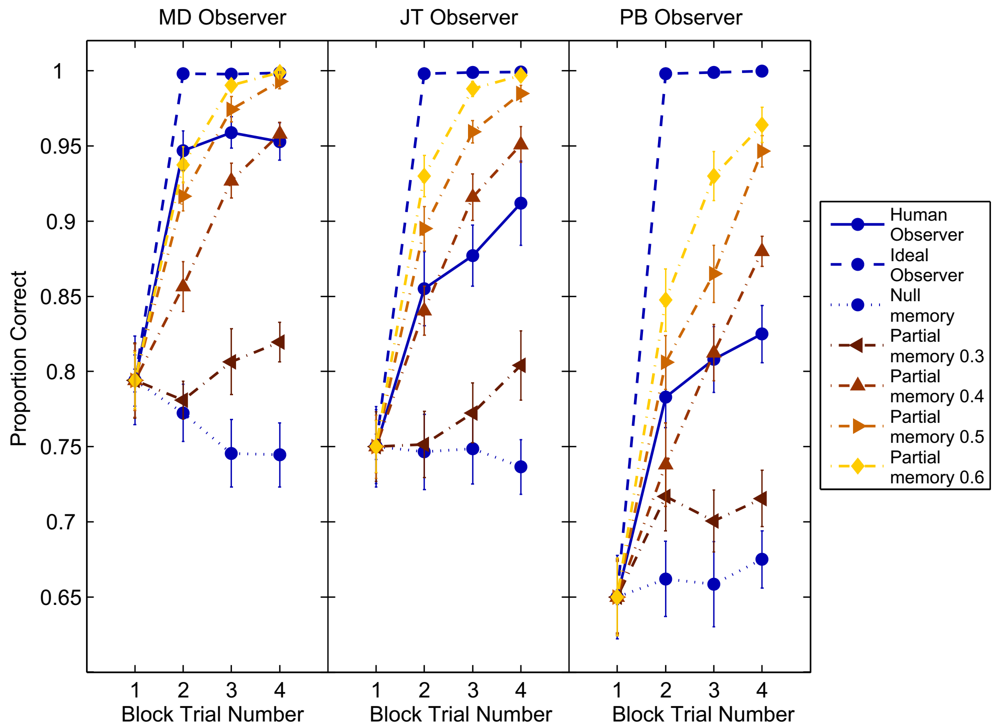

Fig. 9. Comparison of human performance and two models: ideal observer and the sub-optimal model with partial memory for the feedback indicated target location. Simulation results show that the partial memory model variations in the parameter *fm* reduces the amount of learning. The model cannot capture the shape of the performance curve for human observers. Each panel shows the results of one observer for the 10 dots experimental situation.

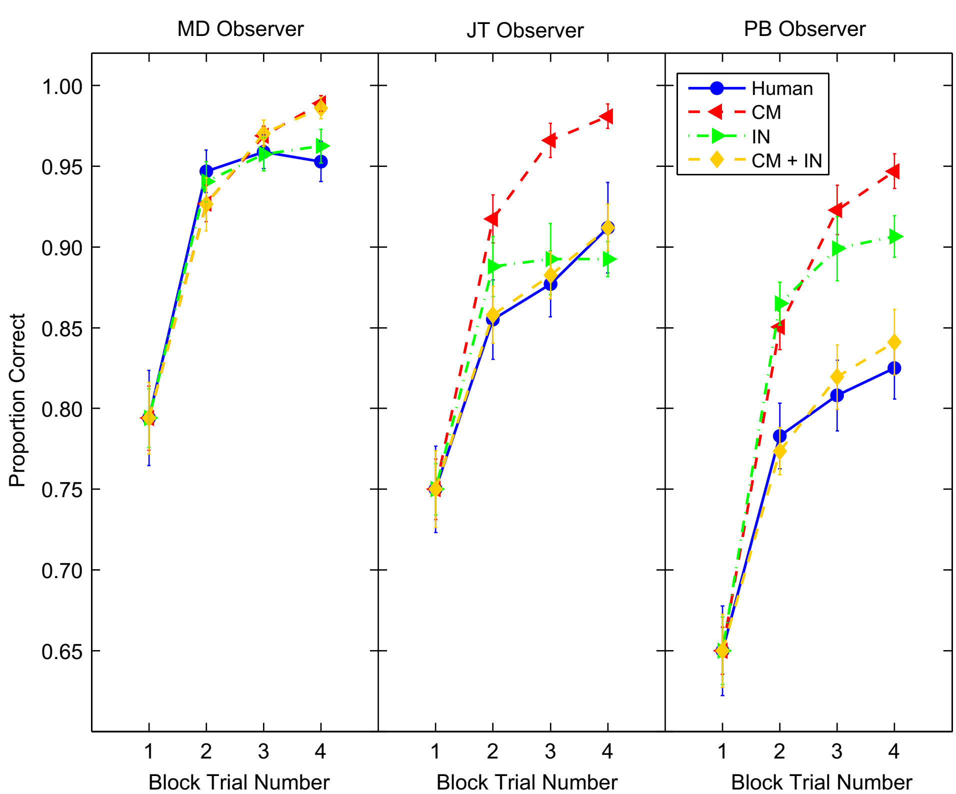

Fig. 10. Comparison among performance obtained by human observers and three sub-optimal models: cautious updating model (CM), noisy memory representation model (IN), and the combinations of both models (CM + IN). Simulations show that for observers JT and PB the combination of CM and IN fits very well the human data. For observer MD, on the other hand, the best fit was achieved with IN model.

response of the model was very sensitive to this parameter so, multiple simulations with small variations of noise were necessary to obtain an acceptable result (see Fig. 11).

5. Discussion

We have presented a series of experiments examining how humans learn about stimulus' motion parameters along blocks of few trials. Each block consisted in four trials in which subjects performed a speed discrimination task and a final direction identification response. Human learning was analyzed in the context of a Bayesian ideal observer, which was specifically derived for the OPL paradigm (Eckstein et al., 2004). This ideal observer uses the feedback provided after each trial to update prior probabilities and thus, improve performance. Psychophysical results show a significant performance improvement during the four trial blocks, for the three observers and both experimental situations (10 and 20 dots per patch) suggesting that the process of learning the relevant motion direction of the signal can rapidly have important effects on improving human judgments about speed (e.g., learning the relevant direction of motion of a ball moving in a crowd of players can improve estimates of its speed). The human performance improvements in speed identification were very large (15–30%). This improvement in proportion correct is greater than the improvements measured over four trials for localization of elongated Gaussian (Eckstein et al., 2004; Abbey et al., 2008) and for face identification (10–15%). This finding could be interpreted as suggesting that humans are better at learning motion than contrast defined targets or faces. However, comparison with the optimal Bayesian learner with the efficiency metric shows that the human learning was much less than that of the ideal observer (Fig. 8). In fact, the present task shows the highest drops in absolute in efficiency (factor of 3–7) as a function of learning trial when compared to previously studied detection of simple targets (elongated Gaussians; Abbey et al., 2008; Eckstein et al., 2004) and faces embedded in luminance noise (Peterson et al., 2009). Thus, although the improvements in proportion correct are highest for the motion dot stimuli compared to other stimuli (faces and simple contrast defined targets) humans are the most sub-optimal relative to the ideal learner. This highlights the importance of using ideal observer to quantify the amount of information to be learned for a specific stimulus and shows how comparisons of proportion correct across tasks can lead to misleading conclusions.

How can we explain the low efficiency of human learning for our motion display? One possible explanation is that humans are more inefficient at extracting the motion direction information from the patch relative to their extraction of the speed information. This might be caused by the fact that observers' make the speed discrimination during or immediately after the stimulus is presented. However, they might wait for the location feedback to assess the posterior probability of the motion directions relying on a degraded memory representation of the extinct stimulus which might severely reduce efficiency. Here we evaluated three different effects of memory on the updating of weights for each motion direction. The 1st model considered an inefficient updating based on sensory evidence from all four patches rather than optimal strategy which is to use the sensory evidence at the patch/location indicated by the feedback. The 2nd model considered that observers only updated the weights when they correctly identified the speed which is consistent with previous studies (Eckstein et al., 2004; Eckstein, Peterson, Pham, & Droll, 2009). The 3rd model included a noisier representation.

Proportion of correct responses were higher for 20 than for 10 dots but we did not find a systematic effect of the number of dots on the amount of learning. The improvement in performance for human observers increases approximately from 15% to 30% and, for the ideal observer, from near 30% to 60%. Interestingly, absolute and learning efficiencies are higher for 10 than for 20 dots. This

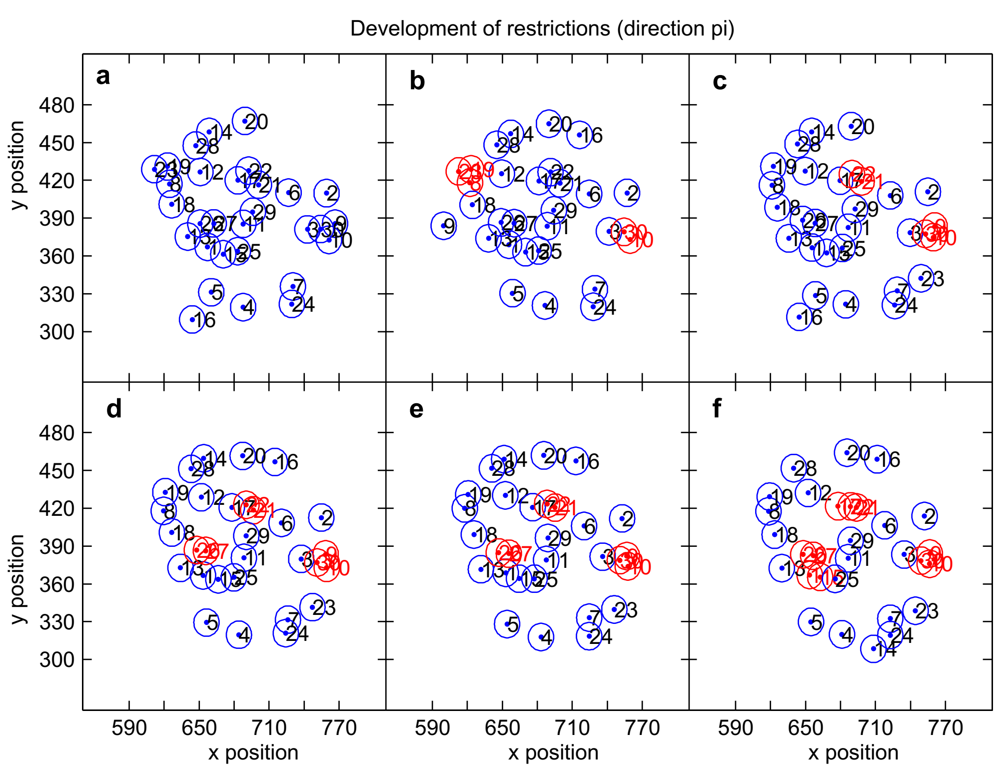

Fig. 11. Dots position in pixels of six successive frames (panels *a–f*) of the right patch. The circles indicate the restriction adopted to minimize the effect of correspondence noise. Red dots are those that do not meet the restriction.

finding is explained by the sub-optimal integration of the additional information as the number of dots increase. The greater visual information given by the increase of number of dots from 10 to 20 results in a smaller increase of performance for the human observer relative to the optimal algorithm. In other words, the ideal observer is more capable in taking advantage of the increase of information than the human observer. A similar lowering absolute efficiency of human performance has been reported when increasing the spatial and/or temporal extent of a signal in luminance noise for detection and discrimination (Kersten, 1984; Eckstein, Whiting, & Thomas, 1996; Tyler & Chen, 2006).

Previous psychophysical experiments (Vaina et al., 1995) reported that humans' performance, in a motion direction discrimination task, can improve in a single session over less than 300 trials, which is considered as rapid learning respect to that occurring with training over days or weeks. Because in their study this process took place in a few minutes, the authors were able to study the neural substrates by using fMRI (Vaina, Belliveau, des Roziers, & Zeffiro, 1998). They hypothesized that learning should be represented by changes in neuronal responses of area MT, as had been proposed by Ball and Sekuler (1982, 1987) based on the fact that learning was direction specific and showed a large amount of binocular transfer. Moreover, Zohary, Celebrini, Britten, and Newsome (1994) had shown that behavioral improvement in motion direction discrimination was accompanied by an increase in neuronal sensitivity at MT. Vaina and collaborators (1998) found that, consistently with their hypothesis, the improvement of psychophysical performance was directly correlated with an expansion of the area of cortical activation in the MT complex with a concurrent reduction of activity in other extrastriate areas.

In the present study we hypothesized that the increase of performance we find in our experiment responds to a rapid reduction of uncertainty produced by an attention optimization implemented through a more optimal integration of speed information across motion directions. We suggest that this process of favoring one motion direction over others could be explicitly implemented in area MT through feedback processes from higher areas and through neural implementations that approximate the optimal Bayesian weighting (Eckstein et al., 2009). The plausibility of the involvement of MT and other attention related areas is supported by Vaina and collaborators (1998) who found that the increase of activity in the MT complex obtained after 300 trials of training correlated with an important reduction of activity in areas of the cerebellum involved in visual-attention tasks (Allen, Buxton, Wong, & Courchesne, 1997) and in the superior colliculus, which is involved in the modulation of motion-related activity by attentional load (Rees, Frith, & Lavie, 1997). However, the specific involvement of MT in the rapid learning observed in our tasks should be tested directly. In fact, our experimental paradigm is well suited to be explored by mean of fMRI studies and/or multielectrode array recordings, which would help to shed light on the mechanisms underlying these rapid processes of perceptual learning.

To summarize our results demonstrate that the effect of rapid learning via uncertainty reduction seems to be ubiquitous in the human visual system and generalizes to a number of feature dimensions and stimuli types including motion, localization of contrast defined targets and faces (Abbey et al., 2008; Eckstein et al., 2004; Peterson et al., 2009).

Appendix A

A.1. Circular values

Because in these experiments the directions lie on the circle of 2π radians, we must be careful in the definition of directions to be accounted by the ideal observer. The following example illustrates this problem: when the algorithm evaluates the direction of a dot and computes the probability to observe that direction given each distribution, there are two cases in which two equivalent distributions, in term of angles, give completely different probability densities. For example, if one considers the patch whose mean direction is $\pi/2$, the distribution with mean 0 (radians) is much nearer than the equivalent distribution with mean 2π and thus, they will give completely different probability densities. To solve this problem, the ideal observer computes the conditional densities as the sum of both equivalent distributions.

A.2. Correspondence noise

Correspondence noise is the uncertainty introduced by the fact that any dot in a given frame may come from multiple dots of the previous frame (Barlow & Tripathy, 1997; Williams & Sekuler, 1984). In displays containing moving random-dot patterns, with added noise to both speed and direction, such is the case of this experiment, there is a finite probability that a dot in a given frame come from any dot from the previous frame. Therefore, to quantify the correspondence noise in such displays one should consider all these possibilities for each dot, which would be extremely costly in terms of computation. To minimize this problem we restricted the position of dots in the first frame such that the distance between two dots is always larger than 20 times the standard deviation of the dots displacement per frame. With this restriction, we avoid the most probable mismatches since displacements larger than 20 standard deviations may be considered as marginal. However, this procedure does not guarantee that in the subsequent frames some dots fail to meet the restriction, due to the distribution of directions and speeds. Fig. 10 shows, as an example, the dots position and the fulfillment of the restriction in six successive frames of a sampled stimulus. The red dots are those that do not meet the restriction because they fall into the circle of other dot and thus, are potential sources of noise. Of course, there are no red dots in the first frame.

We now can estimate the correspondence noise in our stimuli by analyzing only the dots that do not meet the restriction. We consider a correspondence fail that case in which the dot producing the displacement with higher probability density is not the correspondent dot. We analyzed ten stimuli and obtained a mean level of noise of 0.001389 with 0.003526 of standard deviation. We consider these values are small enough to disregard the correspondence noise in the simulations of the ideal observer.

References

Abbey, C. K., Pham, B. T., Shimozaki, S. S., & Eckstein, M. P. (2008). Contrast and stimulus information effects in rapid learning of a visual task. *Journal of Vision, 8*, 1–14.

Allen, G., Buxton, R. B., Wong, E. C., & Courchesne, E. (1997). Attentional activation of the cerebellum independent of motor involvement. *Science, 275*(5308), 1940–1943.

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science, 218*(4573), 697–698.

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Visual Research, 27*, 953–965.

Barlow, H. B. (1980). The absolute efficiency of perceptual decisions. *Philosophical Transactions of the Royal Society of London, B, 290*, 71–82.

Barlow, H. B., & Tripathy, S. P. (1997). Correspondence noise and signal pooling in the detection of coherent visual motion. *The Journal of Neuroscience, 17*(20), 7954–7966.

Beard, B. B., & Ahumanda, A. J. (1999). Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. *Journal of the Optical Society of America A, 16*(3), 755–763.

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

Curran, W., & Braddick, O. J. (2000). Speed and direction of locally-paired dot patterns. *Vision Research, 40*, 2115–2124.

Dosher, B. A., & Lu, Z. L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences of the USA, 95*(23), 13988–13993.

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. *Journal of Vision, 4*, 1006–1019.
Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science, 17*(11), 973–980.
Eckstein, M. P., Peterson, M. F., Pham, B. T., & Droll, J. A. (2009). Statistical decision theory to relate neurons to behavior in the study of covert visual attention. *Vision Research, 49*(10), 1097–1128.
Eckstein, M. P., Shimozaki, S. S., & Abbey, C. K. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *Journal of Vision, 2*(1), 25–45.
Eckstein, M. P., Whiting, J. S., & Thomas, J. P. (1996). Detection and contrast discrimination of moving signals in uncorrelated Gaussian noise. *Proceedings of the SPIE, 2712*, 9–25 [Medical Imaging 1996: Image Perception, Harold L. Kundel; Ed].
Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research, 33*, 397–412.
Fine, I., & Jacobs, R. A. (2000). Perceptual learning for a pattern discrimination task. *Vision Research, 40*, 3209–3230.
Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: A review. *Journal of Vision, 2*, 190–203.
Geisler, W. S. (2003). Ideal observer analysis. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences*. Boston: MIT Press.
Gibson, E. J. (2000). *An ecological approach to perceptual learning and development*. Oxford, NY: Oxford University Press.
Gold, J. M. (2003). Dynamic classification images reveal the effects of perceptual learning in a hyperacuity task. *Journal of Vision, 3*(9), 162a.
Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature, 402*(6758), 176–178.
Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 582–612.
Hulbert, A. (2000). Visual perception: Learning to see through noise. *Current Biology, 10*, 231–233.
Jacobs, R. A. (2009). Adaptive precision pooling of model neuron activities predicts the efficiency of human visual learning. *Journal of Vision, 9*(4), 1–15.
Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature, 16*, 250–252.
Kersten, D. (1984). Spatial summation in visual noise. *Vision Research, 24*(12), 1977–1990.
Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology, 13*(2), 150–158.
Kinchla, R. A., Chen, Z., & Evert, D. (1995). Precue effects in visual search: Data or resource limited? *Perception & Psychophysics, 57*, 441–450.
Liu, Z., Kersten, D., & Knill, D. C. (1999). Dissociating stimulus information from internal representation – A case study in object recognition. *Vision Research, 39*, 603–612.
Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research, 35*, 549–568.
Matthews, N., Liu, Z., Geesaman, B. J., & Qian, N. (1999). Perceptual learning on orientation and direction discrimination. *Vision Research, 39*, 3692–3701.
McKee, S. P., & Westheimer, G. (1978). Improvement in vernier acuity with practice. *Perception & Psychophysics, 24*, 258–262.
Murray, R. F., Sekuler, A. B., & Bennett, P. J. (2003). A linear cue combination framework for understanding selective attention. *Journal of Vision, 3*(2), 116–145.
Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.
Peterson, M. F., Abbey, C. K., & Eckstein, M. P. (2009). The surprisingly high human efficiency at learning to recognize faces. *Vision Research, 49*(3), 301–314.
Rees, G., Frith, C. D., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science, 278*(5343), 1616–1619.
Saarinen, J., & Levi, D. M. (1995). Perceptual learning in vernier acuity: What is learned? *Vision Research, 35*, 519–527.
Saffell, T., & Matthews, N. (2003). Task-specific perceptual learning on speed and direction discrimination. *Vision Research, 43*, 1365–1374.
Shimozaki, S. S., Eckstein, M. P., & Abbey, C. K. (2003). Comparison of two weighted integration models for the cueing task: Linear and likelihood. *Journal of Vision, 3*(3), 209–229.
Tjan, B. S., & Legge, G. E. (1998). The viewpoint complexity of an object-recognition task. *Vision Research, 38*, 2335–2350.
Tyler, C. W., & Chen, C. C. (2006). Spatial summation of face information. *Journal of Vision, 26*(10), 1117–1125.
Vaina, L. M., Belliveau, J. W., des Roziers, E. B., & Zeffiro, T. A. (1998). Neural systems underlying learning and representation of global motion. *Proceedings of the National Academic of Sciences of the USA, 13*, 12657–12662.
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.
Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research, 24*(1), 55–62.

Further reading

Bruyn, B. D., & Orban, G. A. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Research, 28*(12), 1323–1335.
Crist, R. E., Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position, and context. *Journal of Neurophysiology, 78*(6), 2889–2894.
Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
Knill, D. C. (1996). *Perception as Bayesian inference*. Boston: Cambridge University Press.
Wallace, J. M., & Mamassian, P. (2003). The efficiency of speed discrimination for coherent and transparent motion. *Vision Research, 43*, 2795–2810.
Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63*(8), 1293–1313.