

# Dissociable perceptual-learning mechanisms revealed by diffusion-model analysis

Alexander A. Petrov · Nicholas M. Van Horn · Roger Ratcliff

Published online: 11 March 2011  
© Psychonomic Society, Inc. 2011

**Abstract** Performance on perceptual tasks improves with practice. Most theories address only accuracy data and tacitly assume that perceptual learning is a monolithic phenomenon. The present study pioneers the use of response time distributions in perceptual learning research. The 27 observers practiced a visual motion-direction discrimination task with filtered-noise textures for four sessions with feedback. Session 5 tested whether the learning effects transferred to the orthogonal direction. The diffusion model (Ratcliff, *Psychological Review*, 85, 59–108, 1978) achieved good fits to the individual response time distributions from each session and identified two distinct learning mechanisms with markedly different specificities. A stimulus-specific increase in the drift-rate parameter indicated improved sensory input to the decision process, and a stimulus-general decrease in nondecision time variability suggested improved timing of the decision process onset relative to stimulus onset (which was preceded by a beep). A traditional  $d'$  analysis would miss the latter effect, but the diffusion-model analysis identified it in the response time data.

Performance on perceptual tasks improves with practice. This *perceptual learning* occurs in all sensory modalities (Fahle & Poggio, 2002) and is of great theoretical interest

**Electronic supplementary material** The online version of this article (doi:10.3758/s13423-011-0079-8) contains supplementary material, which is available to authorized users.

A. A. Petrov (✉) · N. M. Van Horn · R. Ratcliff  
Department of Psychology, Ohio State University,  
200B Lazenby Hall,  
Columbus, OH 43210, USA  
e-mail: apetrov@alexpetrov.com

(Lu, Yu, Watanabe, Sagi, & Levi, 2009) and practical importance (Polat, 2009). However, its mechanisms are still poorly understood. The signature property of perceptual learning is its *stimulus specificity*: The improvement is (partially) restricted to stimuli similar to those used in training (see, e.g., Ahissar & Hochstein, 1997). This indicates that (part of) the neural substrate of the learning effect resides in the early stages of the sensory processing pathway (Karni & Sagi, 1991), which may involve changes in the early sensory areas (see Gilbert, Sigman, & Crist, 2001, for a review) and/or the read-out connections to decision areas (Doshier & Lu, 1998; Law & Gold, 2008; Petrov, Doshier, & Lu, 2005).

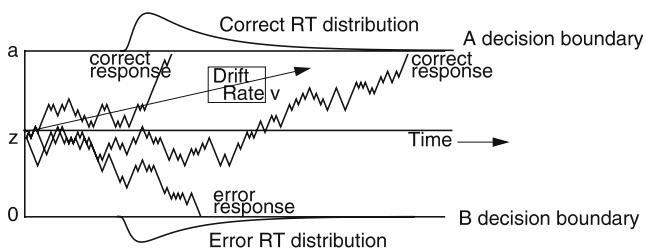
Most perceptual learning studies use accuracy (or, conversely, sensitivity) as their dependent variable (Fahle & Poggio, 2002). Response times (RTs) are typically ignored, or sometimes the mean RTs are analyzed but errors are ignored (e.g., Ding, Song, Fan, Qu, & Chen, 2003). Either approach neglects the well-known speed–accuracy trade-off (Pachella, 1974). Also, such analyses use only one data point per block per stimulus type. This restricted empirical base tends to give rise to restrictive theoretical accounts that attribute all learning to one cortical site (e.g., Karni & Sagi, 1991) or to one learning rule operating at task-dependent levels of the processing hierarchy (Ahissar & Hochstein, 2004).

It seems highly unlikely, however, that perceptual learning is a monolithic phenomenon. Rather, even the simplest task engages multiple brain systems, and the overall behavioral improvement arises from multiple contributions. Univariate data tend to obscure this heterogeneity, whereas richer data sets reveal it. For example, dissociable learning mechanisms have been identified using event-related potentials (Ding et al., 2003), fMRI (Li, Mayhew, & Kourtzi, 2009; Vaina, Belliveau, des Roziers,

& Zeffiro, 1998), and external-noise manipulations (Doshier & Lu, 1998).

The present study pioneers the use of RT distributions for studying perceptual learning. A typical RT distribution can be described approximately by five quantiles (Ratcliff, 1979) that divide the probability mass into six bins. This makes 12 bins in total, for correct and error responses. Since the total mass is fixed, there are 11 degrees of freedom per block. Obviously, this carries much more information than does accuracy alone, but it presents an analytic challenge. We use the diffusion model (DM; Ratcliff, 1978) to analyze such data. This is analogous to the use of signal detection theory (Macmillan & Creelman, 2005) to convert hits and false alarms into theoretically motivated estimates of discriminability and bias. Analogously, DM converts hits, false alarms, and RT distribution statistics into estimated parameters of various processing components.

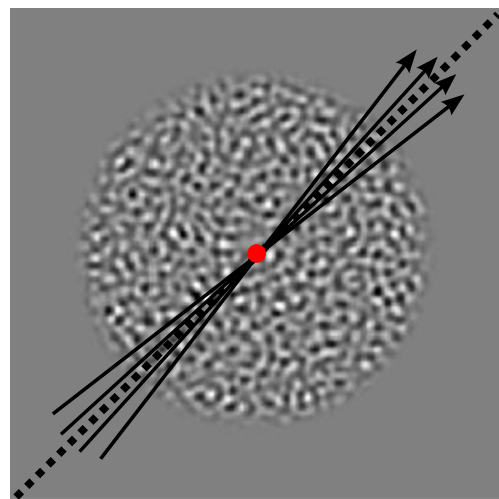
DM characterizes the process of making simple two-choice decisions (see Ratcliff & McKoon, 2008, for a review). The core of the model is a diffusion process that describes the stochastic accumulation of evidence for two competing responses (Fig. 1). The process terminates when the accumulated evidence reaches one of two decision boundaries (or *criteria*). The better the information about the stimulus, the larger the drift rate  $v$  in the correct direction. Due to within-trial variability in evidence accumulation, processes with the same mean drift rate terminate at different times (producing RT distributions) and sometimes at the wrong boundary (producing errors). The model has seven free parameters: mean drift rate  $v$ , across-trial variability  $\eta$  in drift rate, boundary separation  $a$ , mean starting point  $z$  between the boundaries ( $0 < z < a$ ), across-trial range  $s_z$  in starting point, mean nondecision time  $T_{er}$ , and across-trial range  $s_t$  in nondecision time. The first five parameters affect both accuracy and speed in the model. For example, increased drift rate produces higher accuracy and faster RTs, whereas increased boundary separation produces higher accuracy but slower RTs, all



**Fig. 1** Schematic illustration of the decision process in the diffusion model. Three stochastic paths result in a fast correct response, a slow correct response, and an error response. Because of the positive drift rate  $v$ , the correct (upper) boundary is reached more often than the incorrect (lower) boundary. RT, response time;  $a$ , boundary separation parameter;  $z$ , starting point parameter

else being equal. The two nondecision parameters ( $T_{er}$  and  $s_t$ ) affect the RTs only. They describe the combined duration of processes such as stimulus encoding, memory access, and response execution. All parameters are estimated simultaneously by fitting the model to behavioral data. DM is tightly constrained, particularly in experimental designs involving stimuli at multiple difficulty levels. The proportions correct and the shapes of all RT distributions across all difficulty levels must be accounted for with a fixed parameter set within a block. Only the drift rate  $v$  is allowed to vary as a function of difficulty. DM has been tested and validated extensively. For instance, speed-related instructions affect the boundary separation parameter, whereas stimulus manipulations affect the drift rate (see Ratcliff & McKoon, 2008, for a review). However, the effects of practice have been studied relatively little (Dutilh, Krypotos, & Wagenmakers, *in press*; Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009; Ratcliff, Thapar, & McKoon, 2006), and the effects of perceptual learning are largely unknown.

Various perceptual learning mechanisms can be identified using the DM framework. First and foremost, the drift rates are expected to improve with practice (Dutilh et al., 2009; Ratcliff et al., 2006). The drift rates in perceptual tasks measure the quality of the sensory input to the decision process. Most theories of perceptual learning (e.g., Ahissar & Hochstein, 2004; Karni & Sagi, 1991; Petrov et al., 2005) predict a drift-rate increase. This increase affects both accuracy and RT. The diffusion analysis captures both effects in a single measure, and thus can detect weaker learning effects. It can also estimate the variability attributable to speed-accuracy trade-offs across observers and across sessions. Removing this variability from the

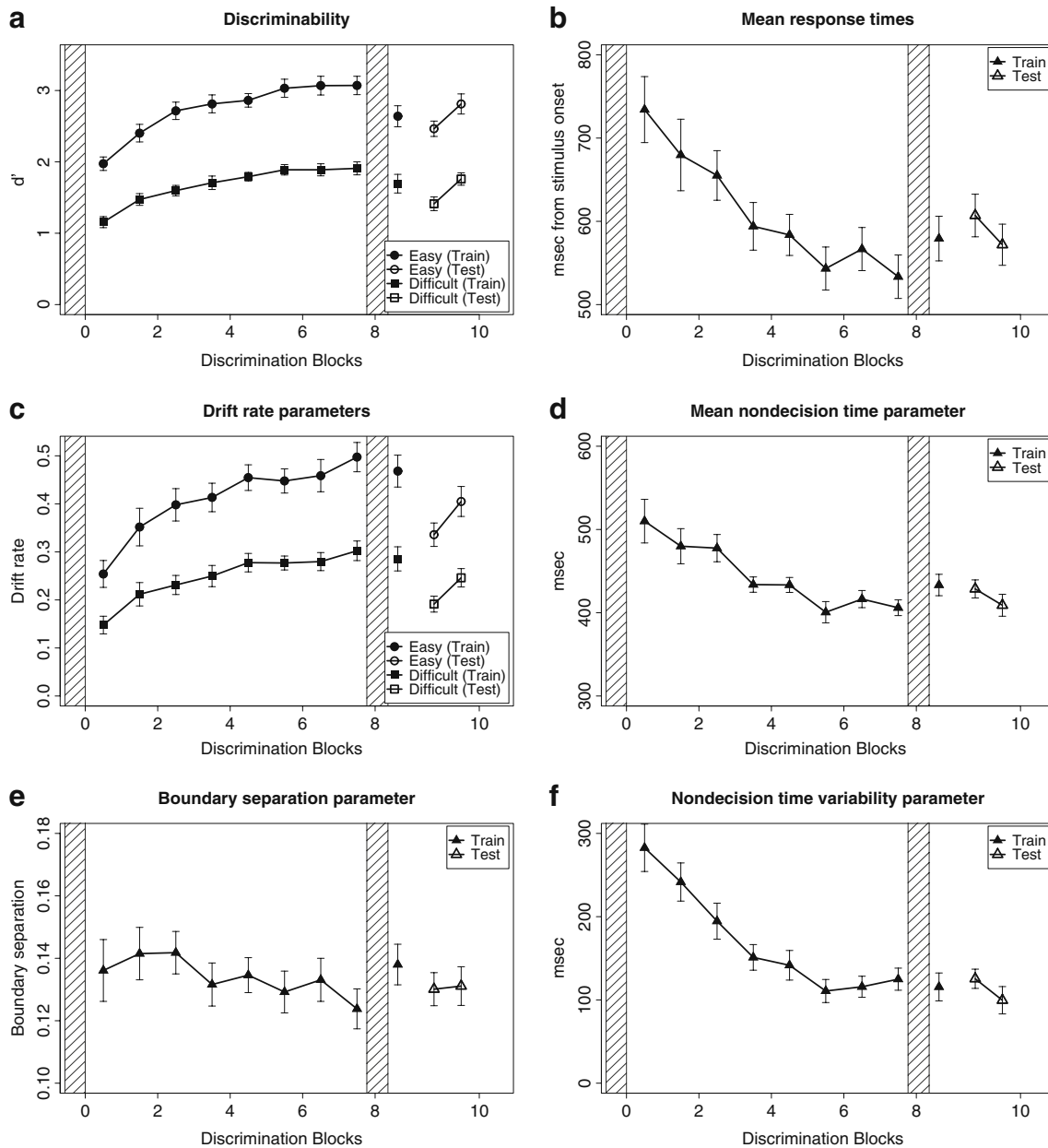


**Fig. 2** The stimuli were filtered-noise textures moving behind a circular aperture. The direction of motion could take four possible values relative to an implicit reference direction (dashed line). The angles are exaggerated for visibility

error term improves the power of the analysis still further (Liu & Watanabe, 2010).

The DM framework also allows us to study the temporal aspects of perceptual learning. Our hypothesis is that observers may learn to deploy attention during the critical period of the trial sequence. Attention plays an important role in perceptual learning (Ahissar & Hochstein, 2002) and in learning more generally (Kruschke, 2003). Though relatively neglected in the literature, the temporal aspects

of attention are as important as its spatial aspects (Large & Jones, 1999). In particular, the decision mechanism must be timed relative to the stimulus onset (Purcell et al., 2010; Ratcliff & Smith, 2010). If triggered before sensory evidence becomes available, the diffusion process merely accumulates noise. If triggered too late, valuable evidence can be lost. Therefore, one way to improve performance is to calibrate the onset of the decision process. To test this *synchronization hypothesis*, we conducted an experiment in



**Fig. 3** Learning profiles for the group-averaged discriminability (a) and mean response times (b) in the raw data, and for various parameters of the diffusion model (c–f). The observers practiced motion-direction discrimination for eight blocks (black symbols) and

then were tested on the same task with motion in the orthogonal direction (open symbols). The error bars are 90% within-subjects confidence intervals. Shaded areas mark two additional sessions of motion aftereffect measurements.

**Table 1** Descriptive statistics of the discriminability  $d'$  for easy and difficult stimulus pairs and for all diffusion-model parameters

Variable	Learning Index LI	Specificity Index SI	Linear Trend		Quadratic	
			$F$	$\eta_p^2$	$F$	$\eta_p^2$
Easy discriminability $d_1'$	.55±.08	.55±.09	152***	.74	24.6***	.30
Difficult discriminability $d_2'$	.65±.10	.66±.12	165***	.75	23.6***	.29
Mean discriminability $d'$	.55±.08	.60±.10	202***	.79	29.7***	.35
Mean reaction time RT	-.27±.03	.37±.08	251***	.82	19.0***	.25
Easy drift rate $v_1$	0.96±0.23	.66±.10	91***	.62	10.0**	.14
Difficult drift rate $v_2$	1.05±0.25	.72±.09	95***	.64	10.2**	.14
Mean drift rate $v$	0.99±0.23	.68±.09	98***	.64	10.7**	.15
Mean nondecision time $T_{cr}$	-.20±.04	.22±.10	62***	.53	4.32*	.06
Nondecision time range $s_r$	-.56±.06	.00±.08	161***	.75	28.0***	.33
Boundary separation $a$	-.09±.07	n/a	7.32**	.10	0.68	.00
Drift variability $\eta$	.37±.21	n/a	6.00*	.08	2.04	.02
Starting point range $s_z$	-.03±.25	n/a	0.37	.00	0.20	.00

The learning and specificity indices were calculated from group-averaged data ( $\pm 80\%$  bootstrap confidence interval; see note 2 for details). The trend analyses were performed on the individual data using orthogonal contrasts (linear and quadratic). The error term had  $df = 182$  in all ANOVAs.  $\eta_p^2$  = partial effect size = variance of the contrast relative to itself plus the error. \*  $p < .100$ . \*\*  $p < .010$ . \*\*\*  $p < .001$

which a beep reliably preceded the stimulus onset. The critical prediction was that the variability of nondecision times would be high at first and decrease significantly with practice. Moreover, this decrease should transfer across stimuli as long as the temporal structure remains the same on each trial. In contrast, the improvement in drift rates was expected to be partially stimulus specific. The experiment included a transfer test to assess the specificity of learning. The task was a visual motion-direction discrimination.

## Method

### Participants

A total of 27 university students with normal or corrected-to-normal vision were paid \$6/h plus a bonus contingent on their accuracy.

### Stimuli and apparatus

The stimuli were filtered-noise textures moving behind a circular aperture (see Fig. 2; diameter  $10^\circ$ , speed  $12^\circ/s$ ). The filter had a Gaussian cross-section along the frequency axis in the Fourier domain (peak frequency 3 cycles/deg at all orientations, full width at half height 4 octaves). On each trial, the filter was applied to a fresh sample of independent, identically distributed Gaussian noise. All stimuli were generated in MATLAB and presented on a 21-in. NEC Accusync 120 color CRT (96 frames/s, mean luminance  $16.6 \text{ cd/m}^2$ , chinrest at 93 cm, free viewing in a darkened room). Each trial began with a brief beep. The texture appeared 500 ms later, moved for 397 ms, and then disappeared. The beep onset always preceded the texture onset by exactly 500 ms, and thus could serve as a reliable attentional cue.

### Task and procedure

The direction discrimination task was defined with respect to a *reference direction*  $\theta$ . The actual motion direction took four possible values:  $(\theta - 3.5)$ ,  $(\theta - 2)$ ,  $(\theta + 2)$ , and  $(\theta + 3.5)$  degrees from vertical. Each block randomly presented 120 stimuli of each kind. The instructions designated the first two as “counterclockwise” and the other two as “clockwise.” The observers pressed a key with their left hand to respond “counterclockwise,” and another key with their right hand to respond “clockwise.” We used four stimuli in a binary task in order to prevent same–different comparisons with the previous trial and to constrain the DM.

Since the task was quite monotonous, many students from our participant pool tended to sacrifice accuracy for speed. To prevent guessing and keep the observers engaged, the procedure rewarded accuracy and penalized excessively fast RTs. The reward for each correct response was a bonus point. The penalty for each error was the loss of a bonus point, an unpleasant beep, and the addition of 250 ms to the 800-ms intertrial interval. The cumulative bonus was displayed prominently at all times. The penalty for excessively fast ( $< 250$  ms) RTs was a “slow down” message that forced the participant to wait for 1,500 ms. RTs between 250 and 500 ms incurred a silent penalty of  $2^*$  (500–RT) ms. Thus, the fastest way to complete a trial was to produce a correct response in exactly 500 ms.

Each participant completed a total of 4,800 discrimination trials in 10 blocks across five sessions. Two additional sessions—before Block 1 and after Block 8—measured the motion aftereffect (MAE; Mather, Verstraten, & Anstis, 1998) in the trained, test, and two control directions.<sup>1</sup> Fourteen participants trained with  $\theta = -50^\circ$  on Blocks 1–8 and on a

<sup>1</sup> The MAE sessions do not affect the interpretation of the present results. See the online supplement for details.

“mini-block” consisting of the first 120 discrimination trials on the last session. These participants then tested with  $\theta = +40^\circ$  on Blocks 9 and 10. The other 13 participants followed the same schedule but trained on  $\theta = +40^\circ$  and tested on  $\theta = -50^\circ$ .

### Data analysis

The data for clockwise and counterclockwise stimuli were pooled because this distinction had no statistically significant effects. The discriminability ( $d'$ ) was calculated for the easy ( $\Delta = 7^\circ$ ) and difficult ( $\Delta = 4^\circ$ ) pairs in each block. Seven DM parameters (easy  $v$ , difficult  $v$ ,  $a$ ,  $T_{\text{crs}}$ ,  $s_t$ ,  $\eta$ , and  $s_z$ ;  $z = a/2$ ) were estimated for each observer in each block. An iterative algorithm minimized the  $\chi^2$  discrepancy between the predicted and observed quantile RTs (Ratcliff & Tuerlinckx, 2002).

We used trend analysis to test the statistical significance of the learning effects during Blocks 1–8. We also calculated two quantitative indices: The *learning index*  $LI = (X_8 - X_1)/X_1$  for a variable  $X$  quantified the improvement by the end of training relative to the initial performance (Fine & Jacobs, 2002). The *specificity index*  $SI = (X_8 - X_9)/(X_8 - X_1)$  quantified the disruption caused by the switch to the orthogonal direction in Test Block 9 (Ahissar & Hochstein, 1997).

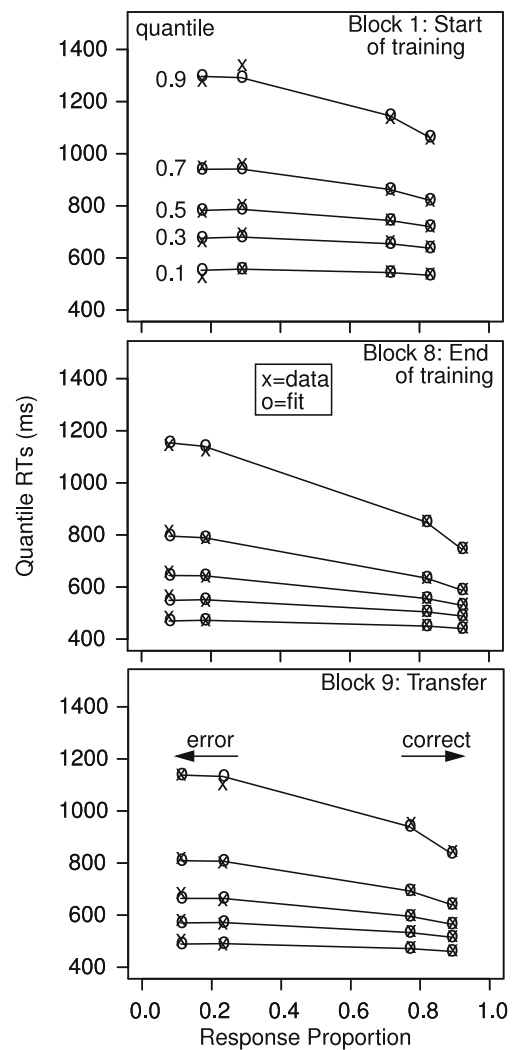
### Results

Both discriminability and mean RT improved with practice (Fig. 3) and showed highly significant linear and quadratic trends (Table 1). The  $d'$  profiles for easy and difficult discriminations were approximately proportional to each other ( $d'_{\text{ez}} \approx kd'_{\text{diff}}$  with  $k = 1.65 \pm 0.13$ ), in agreement with published data (Petrov et al., 2005). The learning effects were partially specific to the trained reference direction, although the degree of specificity differed significantly for the two dependent measures. The specificity index was<sup>2</sup>  $SI = .60 \pm .10$  for  $d'$  and  $.37 \pm .08$  for the mean RT.

The DM achieved good fits, evident in the *quantile probability plots* in Fig. 4 and the scatterplots in Fig. 5. The former show the proportions of correct and error responses (on the x-axis) and the corresponding RT distributions (summarized by the .1, .3, .5, .7, and .9 quantiles on the y-axis). The model (circles) tracks the data ( $\times$ 's) well.<sup>3</sup> The

<sup>2</sup> All indices throughout the text and in Table 1 are reported as  $I \pm CI$ , where  $I$  is the index calculated from the group-averaged data in Figure 3, and  $\pm CI$  is the 80% bootstrap confidence interval. We estimated the variance of the group-level indices by resampling the participants with replacement into 1,000 “groups” and repeating the calculation for each group.

<sup>3</sup> The quantitative measure ( $\chi^2$ ) of goodness of fit confirms this. See the online supplement for details.

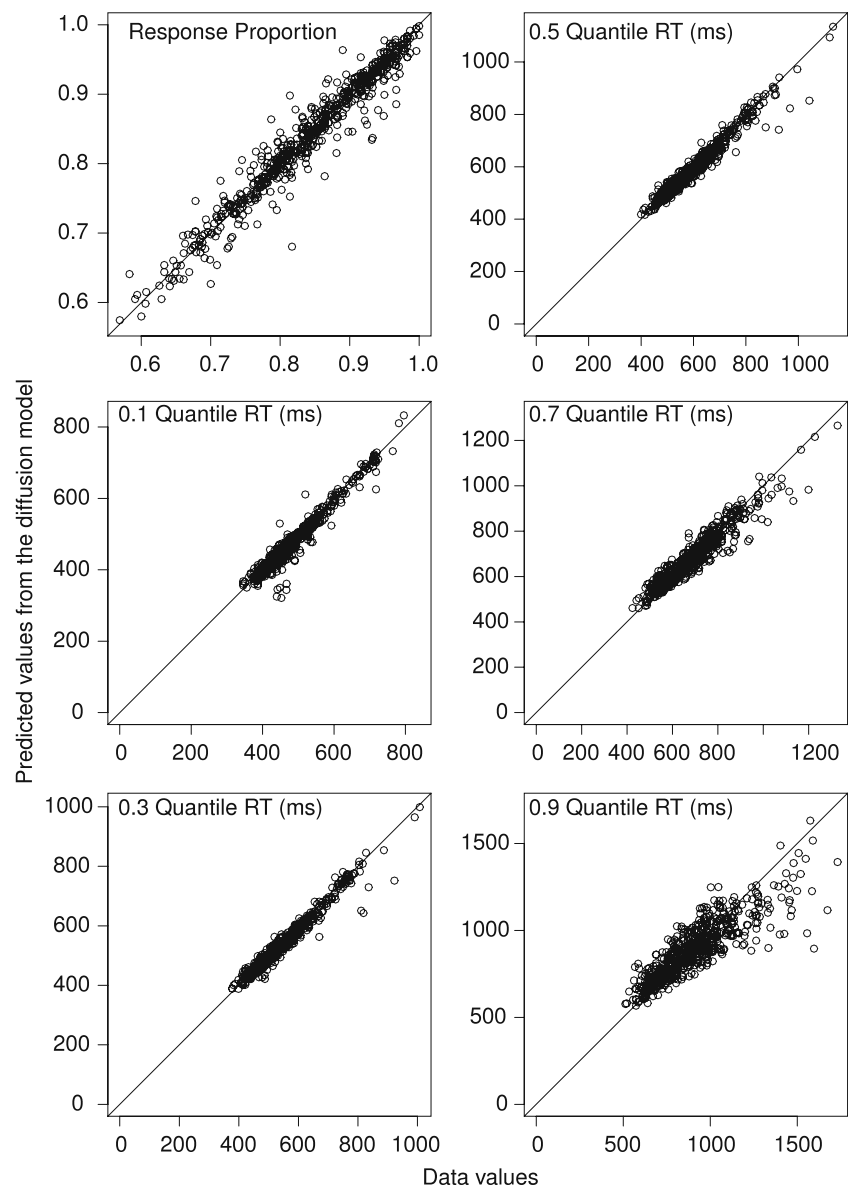


**Fig. 4** Quantile probability plots illustrating the wealth of data and the quality of the fit. Each panel has 22 empirical degrees of freedom: the proportions of errors and correct responses for the easy and difficult discriminations (plotted on the x-axis) and the .1, .3, .5, .7, and .9 quantiles of the corresponding response time distributions (stacked vertically on the y-axis). For example, the x-coordinate of the leftmost, bottommost data point in the top panel indicates the initial error rate (.18) for easy stimuli. The y-coordinate indicates the leading edge (.1 quantile  $\approx 530$  ms) of the corresponding RT distribution. After 4 days of training (middle panel), the performance improves on both measures (.08 rate and 480 ms, respectively). This illustration is based on group-averaged data; the analyses in the text (and the predictions in Fig. 5) are based on model fits to individual data

scatterplots show that the model can reconstruct the data for each individual on each block to a good approximation. The quality of the fit, coupled with past research (Ratcliff & McKoon, 2008) validating the DM in conditions similar to ours, suggests that the DM parameters offer a concise characterization of the underlying cognitive processes.

There were statistically significant learning effects for all DM parameters except the starting point variability  $s_z$  (Table 1). The twofold improvement in drift rate (Fig. 3c)

**Fig. 5** Scatterplots illustrating the quality of the fit to individual data. The diffusion model was fit separately in each block (297 fits = 27 observers  $\times$  11 blocks). Each panel contains 594 points (= 297 $\times$ 2 difficulty levels). RT, response time



indicates that the quality of the sensory input to the decision process increases with practice. The learning index for the drift rate  $\nu$  ( $LI = 0.99 \pm 0.23$ ) was significantly<sup>4</sup> higher than the  $d'$  learning index ( $.55 \pm .08$ ). This is because  $\nu$  reflects learning in both accuracy and speed. The improvement was largely (but not entirely) specific to the trained reference direction ( $SI = .68 \pm .09$ ).

The parameters describing the distribution of nondecision times across trials also improved significantly. The mean nondecision time  $T_{er}$  decreased by 20% on average (Fig. 3d). The specificity index for  $T_{er}$  ( $.22 \pm .10$ ) was significantly<sup>5</sup> lower than that for the mean overall RT ( $.37 \pm .08$ ). This is

because the improvement in overall RT stems in part from the stimulus-specific increase in drift rate.

The nondecision variability  $s_t$  is of particular interest. As predicted by the synchronization hypothesis, it was high at first (283 ms during Block 1, Fig. 3f) and decreased steeply to 120 ms by the end of training. Moreover, the improvement transferred fully to the orthogonal direction of motion ( $SI = .00 \pm .08$ ).

There was a small but statistically significant decreasing linear trend in the boundary separation parameter  $a$  (Table 1). This suggests a slight adjustment in the speed–accuracy trade-off. The drift-rate increase apparently offset this adjustment and prevented a drop in accuracy. Finally, there was a marginally significant decrease in the across-trial variability in drift rate ( $\eta$ ) but no significant changes in the variability in starting point ( $s_z$ ). See the [online supplement](#) for details.

<sup>4</sup> Paired-samples bootstrap  $z = 2.8$ ,  $p < .01$ .

<sup>5</sup> Paired-samples bootstrap  $z = 2.5$ ,  $p < .013$ .

## Discussion

This article makes two contributions: methodological and substantive. The methodological one is to demonstrate the applicability of the diffusion model to perceptual learning research. This research typically involves simple two-choice tasks, RTs faster than 1 s, and thousands of trials—precisely the conditions that DM is best suited for. The model accounted for all behavioral data—22 measurements per block in our experiment—with seven parameters. Figure 5 demonstrates that this reduction occurs with little loss of information. Moreover, the DM parameters have theoretically motivated and empirically validated interpretations in terms of component processes (see Ratcliff & McKoon, 2008, for a review). Neurophysiological correlates of several such processes have been found, narrowing the gap between brain and behavior (see Gold & Shadlen, 2007, for a review).

The DM analysis confers several advantages, all of which stem from its access to more detailed data. First, the drift-rate parameter  $\nu$  is sensitive to improvements in both accuracy and speed. This produces stronger learning effects, manifested here in the high learning index for  $\nu$ . Second, DM accounts for speed–accuracy trade-offs (Dutilh et al., 2009; Ratcliff et al., 2006). The associated variability can be partialled out of the error term (Liu & Watanabe, 2010). The third and most important advantage is that DM reveals phenomena that cannot be reached by traditional methods. This leads to the substantive contribution of this article.

We identified two distinct learning mechanisms with markedly different specificities. The first mechanism improves the quality of the sensory input to the decision process, and manifests itself in increased drift rates. This improvement is partially stimulus specific and is compatible with most theories of perceptual learning, including representation modification (Gilbert et al., 2001; Karni & Sagi, 1991) and selective reweighting (Doshier & Lu, 1998; Law & Gold, 2008; Petrov et al., 2005). The second mechanism manifests itself in decreased and less variable nondecision times.

As discussed in the introduction, triggering the decision process too early or too late impairs performance. Recent theoretical work has suggested that the diffusion process can be gated (Purcell et al., 2010) or disinhibited (Ratcliff & Smith, 2010) at the time in which usable sensory evidence becomes available at the decision-making areas. One intriguing interpretation of our data is that the observers improved the timing of this internal gating operation. We speculate that during the first session, this timing was good on some trials but too slow on others. This inflated the nondecision time variability and degraded the mean drift rates. Apparently, the slow nondecision times

were eliminated with practice. Under this *synchronization hypothesis*, the nondecision time variability  $s_t$  decreased as the observers learned the temporal relationship between the beep and the stimulus onset. Because this relationship was independent of motion direction, the  $s_t$  decrease transferred fully to new stimuli.

The nondecision time in the DM covers the combined duration of stimulus encoding and response execution. Probably both improved with practice. Although the present data cannot differentiate their relative contributions, it seems unlikely that the speed-up can be attributed entirely to motor factors. The drop in  $T_{er}$  (Fig. 3d) was partially stimulus specific, and the 160-ms drop in  $s_t$  (Fig. 3f) seems too large relative to the duration of simple RTs (Luce, 1986).

We used a combination of bonuses and “slow down” messages to prevent fast guessing. Pilot data indicated that without such incentives, some students from our participant pool tended to respond so quickly that their accuracy was barely above chance. Not surprisingly, there was little improvement with practice. Our procedure minimized this behavior and produced robust learning effects. Still, it must be acknowledged that our results may not generalize well to less motivated observers.

In conclusion, perceptual learning is not a monolithic phenomenon. Two learning mechanisms with different properties seem to be at work in our study and can be dissociated with the aid of the diffusion model.

## References

- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*, 401–406.
- Ahissar, M., & Hochstein, S. (2002). The role of attention in learning simple visual tasks. In M. Fahle & T. Poggio (Eds.), *Perceptual learning* (pp. 253–272). Cambridge: MIT Press.
- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*, 457–464.
- Ding, Y., Song, Y., Fan, S., Qu, Z., & Chen, L. (2003). Specificity and generalization of visual perceptual learning in humans: An event-related potential study. *NeuroReport*, *14*, 587–590.
- Doshier, 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*, *95*, 13988–13993.
- Dutilh, G., Krypotos, A.-M., & Wagenmakers, E.-J. (in press). Task-related vs. stimulus-specific practice: A diffusion model account. *Experimental Psychology*.
- Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review*, *16*, 1026–1036.
- Fahle, M., & Poggio, T. (Eds.). (2002). *Perceptual learning*. Cambridge: MIT Press.
- Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: A review. *Journal of Vision*, *2*, 190–203.

- Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, *31*, 681–697.
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.
- Kami, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences*, *88*, 4966–4970.
- Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, *12*, 171–175.
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*, 119–159.
- Law, C.-T., & Gold, J. I. (2008). Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nature Neuroscience*, *11*, 505–513.
- Li, S., Mayhew, S. D., & Kourtzi, Z. (2009). Learning shapes the representation of behavioral choice in the human brain. *Neuron*, *62*, 441–452.
- Liu, C., & Watanabe, T. (2010). Accounting for speed–accuracy tradeoff in visual perceptual learning [Abstract]. *Journal of Vision*, *10*(7), 1111a.
- Lu, Z.-L., Yu, C., Watanabe, T., Sagi, D., & Levi, D. (2009). Perceptual learning: Functions, mechanisms, and applications. *Vision Research*, *49*, 2531–2534.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). New York: Cambridge University Press.
- Mather, G., Verstraten, F. A. J., & Anstis, S. (Eds.). (1998). *The motion aftereffect: A modern perspective*. Cambridge: MIT Press.
- Pachella, R. G. (1974). The interpretation of reaction time in information-processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). New York: Halsted.
- Petrov, A. A., Doshier, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*, *112*, 715–743.
- Polat, U. (2009). Making perceptual learning practical to improve visual functions. *Vision Research*, *49*, 2566–2573.
- Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, *117*, 1113–1143.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*, 446–461.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.
- Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General*, *139*, 70–94.
- Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging, practice, and perceptual tasks: A diffusion model analysis. *Psychology and Aging*, *21*, 353–371.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*, 438–481.
- 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 Academy of Sciences*, *95*, 12657–12662.