Acquisition of decision making criteria: reward rate ultimately beats accuracy

Fuat Balci · Patrick Simen · Ritwik Niyogi ·
Andrew Saxe · Jessica A. Hughes · Philip Holmes ·
Jonathan D. Cohen

Abstract Speed–accuracy trade-offs strongly influence the rate of reward that can be earned in many decision-making tasks. Previous reports suggest that human participants often adopt suboptimal speed–accuracy trade-offs in single session, two-alternative forced-choice tasks. We investigated whether humans acquired optimal speed–accuracy trade-offs when extensively trained with multiple signal qualities. When performance was characterized in terms of decision time and accuracy, our participants eventually performed nearly optimally in the case of higher signal qualities. Rather than adopting decision criteria that were individually optimal for each signal quality, participants adopted a single threshold that was nearly optimal for most signal qualities. However, setting a single threshold for different coherence conditions resulted in only negligible decrements in the maximum possible reward rate. Finally, we tested two hypotheses regarding the possible sources of suboptimal performance: (1) favoring accuracy over reward rate and (2) misestimating the reward rate due to timing uncertainty. Our findings provide support for both hypotheses, but also for the hypothesis that participants can learn to approach optimality. We find specifically that an accuracy bias dominates early performance, but diminishes greatly with practice. The residual discrepancy between optimal and observed performance can be explained by an adaptive response to uncertainty in time estimation.

Keywords Decision making · Dot motion discrimination · Drift-diffusion · Interval timing · Learning · Optimality · Reward rate maximization · Speed–accuracy trade-off · Two-alternative forced-choice

F. Balci (✉) · P. Simen · R. Niyogi · A. Saxe · J. A. Hughes · P. Holmes · J. D. Cohen
Princeton Neuroscience Institute, Princeton University,
Green Hall, Room: 3N12,
Princeton, NJ 08544, USA
e-mail: fbalci@princeton.edu

P. Holmes
Program in Applied and Computational Mathematics
& Department of Mechanical and Aerospace Engineering,
Princeton University,
Princeton, USA

J. D. Cohen
Department of Psychology, Princeton University,
Princeton, USA

Introduction

Humans and animals routinely face simple perceptual and behavioral decisions. However, no matter how simple the decision, decision-makers usually have limited access to information. Thus, accurate decisions require time for evidence accumulation which leads to a speed–accuracy trade-off: spending less time accumulating evidence produces more errors and faster response times (RTs), whereas spending more time leads to greater accuracy and slower response times (Pachella, 1974; Wickelgren, 1977). We take the optimal balance between speed and accuracy to be the one that maximizes the rate of rewards (equivalently, correct responses) earned in a task (Bogacz, Shea-Brown, Moehlis, Holmes, & Cohen, 2006; Gold & Shadlen, 2002). A previous report showed that in a single session, two-alternative forced-choice (2AFC) task in which correct responses were rewarded with money, the majority of the

human participants accumulated more evidence than was necessary to maximize their reward rate (Bogacz, Hu, Holmes, & Cohen, 2010). Here, we investigated whether human participants can acquire an optimal speed–accuracy trade-off after extensive training in a two-alternative forced-choice task with different levels of signal quality (specifically, a random dot motion discrimination task).

Reward-maximizing decision strategies in fixed session-time, 2AFC tasks need to take account of multiple task parameters that include the signal-to-noise ratio (SNR), any penalty for making errors, and the response to stimulus interval (RSI). Thus, we also examined a possible relationship between decision-making performance and endogenous uncertainty about temporal intervals (i.e. RSI). The optimal speed–accuracy trade-off that maximizes the expected reward rate depends critically on the RSI. Poor estimation of this task parameter may therefore be expected to degrade performance (Bogacz et al., 2006; Zacksenhouse, Holmes, & Bogacz, 2010).

Finally, we extended previous investigations (e.g., Bogacz et al., 2010; Simen et al., 2009) by training participants over many sessions with a wide range of signal-to-noise ratios. This allowed us to characterize the acquisition of reward maximizing decision strategies over time and across a large range of SNRs, including the absence of signal (SNR = 0).

The drift diffusion model (DDM): an optimal model for 2AFC

The sequential probability ratio test (SPRT; Barnard, 1946; Wald, 1947) is an optimal statistical procedure for simple, two-alternative hypothesis testing when the number of sequential data samples is unlimited: it minimizes the number of samples for any given level of accuracy (or error proportion, ER); conversely, it maximizes accuracy for any given number of samples (Wald & Wolfowitz, 1948). Stone (1960) first derived an SPRT-based model of choice reaction time in 2AFC tasks, proposing that decision makers computed the likelihood ratio of the two hypotheses under consideration when sampling a signal, and equating the total sample count with decision time. In the drift-diffusion model (DDM), these discrete samples are replaced with a continuous random variable (Ratcliff, 1978), and some SPRT parameters are made variable to allow greater flexibility in fitting data, especially data with unequal average response times for correct responses and errors (Ratcliff & Rouder, 1998). The DDM assumes that the sensory information is noisy, that the difference between the evidence supporting the two alternatives is the decision variable, that this variable is integrated over time, and that when it crosses one of two decision thresholds—one above and one below the starting point or prior belief

state—the corresponding decision is made. The DDM in its most simplified form, which we refer to as the "pure DDM", is given by the first order stochastic differential equation:

$$dx = A dt + \sigma dW, \; x(0) = 0, \tag{1}$$

where x denotes the difference between the evidence supporting two different alternatives at any given time t and can be interpreted as the current value of the log-likelihood ratio, $A dt$ represents the average increase in x during dt, and dW is white noise, Gaussian distributed with mean 0 and variance $\sigma^2 dt$ (see Bogacz et al., 2006; Ratcliff & McKoon, 2008, for detailed reviews of the DDM).

A decision is made when the decision variable crosses either the upper or lower threshold ($\pm z$) and the first passage time is identified with the decision time, DT. The response time (RT) is the sum of the decision time and a non-decision related latency, T_0 (e.g., reflecting sensory encoding and motor execution times). The clarity of the signal is represented by the drift A. Higher values of A therefore lead to greater accuracy and faster response times for a given threshold value z. Speed–accuracy trade-offs arise in the model because of the threshold parameter z: due to noise, lower thresholds lead to faster but less accurate decisions whereas higher thresholds lead to slower but more accurate decisions. Human accuracy and response times in two-alternative forced-choice tasks have been closely fit by an extended version of the DDM that includes trial-to-trial variability in drift rate A, starting point $x(0)$, and non-decision time T_0 (e.g., Ratcliff & Rouder, 1998; Ratcliff, Van Zandt, & McKoon, 1999). These additional parameters make it possible to fit data with unequal correct and error response times. As we show, however, that the pure DDM provides reasonably good fits, and benefits from extremely simple, analytically tractable predictions regarding reward rate maximization (these can only be estimated for the extended DDM by numerical simulation). Testing these simple predictions is the main objective of this paper.

The expected reward rate (RR) in free-response two-alternative forced-choice tasks is the proportion of correct, rewarded responses divided by the average time between them (Gold & Shadlen, 2002):

$$RR = \frac{1 - ER}{DT + T_0 + RSI}, \tag{2}$$

where ER denotes the fraction of errors (ER). Note that in free-response paradigms with a fixed session time, participants have a limited amount of time to make as many decisions as possible. In such cases, the speed of decisions is as important as their accuracy since the response times determine the number of decisions a participant could make during the fixed session time.

When task parameters are fixed within a block, the pure DDM prescribes a parameter-free optimal performance curve (OPC) that defines the reward rate maximizing decision time as a function of ER (Bogacz et al., 2006):

$$\frac{DT}{D_{tot}} = \left(\frac{1}{ER \log \frac{1-ER}{ER}} + \frac{1}{1-2ER} \right)^{-1}, \quad (3)$$

where $D_{tot} = T_0 + RSI$. The optimal performance curve has a skewed inverse U-shape (see Fig. 1, dashed curve) that defines the reward maximizing, mean-normalized decision-times for different error proportions. At the extremes of error proportion (ER = 0 and ER = 0.5) the derived relation makes intuitive sense: for both very high signal-to-noise ratio (ER ≈ 0) and for conditions in which no signal is present (ER ≈ 0.5), the decision threshold should approach 0 in order to maximize reward rate. However, for 0 < ER < 0.5 participants should integrate the sensory information, and there is an optimum threshold that balances between a greater number of decision opportunities (by responding faster) and a greater proportion of accurate decisions (by integrating longer).

We reiterate this important relationship as we will rely on it throughout the paper: When participants are free to respond at any time, the speed–accuracy trade-off they select is crucial for reward rate maximization. Faster response times decrease the accuracy of decisions but at the same time allow more decision opportunities (i.e. the number of trials in a given period), whereas slower response times increase accuracy of decisions but at the same time reduce the number of decision opportunities. This balance between speed and accuracy is important for reward rate maximization since reward rate in the free-
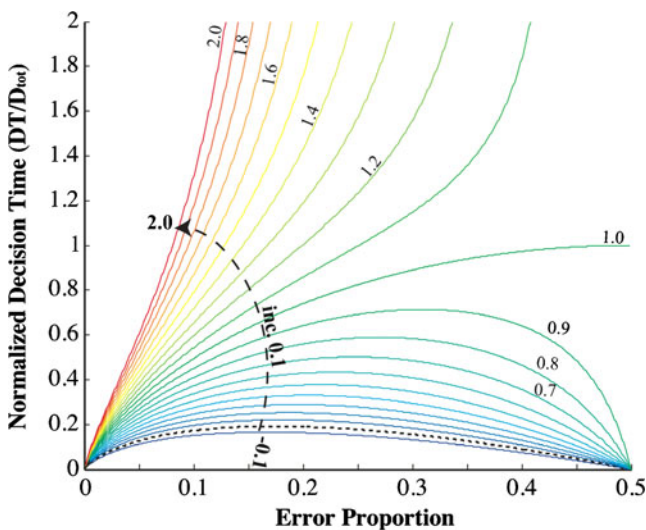

Fig. 1 Optimal performance curves (OPCs) of RR_m parameterized by q in increments of 0.1 (denoted *inc*). *Dashed curve* is the optimal performance curve for $q = 0$, which is equivalent to the optimal performance curve of the pure DDM

response paradigms is $\frac{1-ER}{TrialDuration}$, where *TrialDuration* = $RT + RSI$. This clearly makes the reward rate not only depend on accuracy of decisions but also their speed. For further discussion of this relationship, we refer readers to Bogacz et al. (2006) and Zacksenhouse et al. (2010).

Suboptimal performance in 2AFC: favoring accuracy over reward rate

In contrast to predictions of optimal speed–accuracy trade-off selection, a recent report showed that in single-session 2AFC tasks, the majority of the human participants set their decision thresholds higher than the reward maximizing threshold prescribed by the DDM (Bogacz et al., 2010). This performance pattern is expected under the theory that participants emphasize accuracy over reward (Maddox & Bohil, 1998). To formalize this alternative, Bogacz et al. (2006) defined a modified reward rate function with a penalty for errors,

$$RR_m(q) = \frac{(1 - ER) - qER}{\langle DT \rangle + D_{tot}}, \quad (4)$$

where q is a parameter representing the weight assigned to accuracy relative to reward. Equation 5 defines a family of optimal performance curves for different reward functions RR_m parameterized by q (Fig. 1, solid curves), whose peaks are shifted upward and rightward for increasing values of q (for 0 < q ≤ 1):

$$\frac{DT}{D_{total}} = (1 + q) \left(\frac{\frac{1}{ER} - \frac{q}{1-ER}}{\log \frac{1-ER}{ER}} + \frac{1-q}{1-2ER} \right)^{-1}. \quad (5)$$

Note that, for $q = 0$, Eqs. (4) and (5) reduce to Eqs. (2) and (3). With q as a free parameter, this model fits behavioral data better than the parameter-free DDM, lending support for a model that assumes competition between accuracy and reward (e.g., Holmes et al., 2005; Zacksenhouse et al., 2010)[1]. For q > 1 (penalty for an error exceeds reward for correct response), the optimal strategy is to spend enough time integrating evidence so that error proportion (ER) falls below a critical value, ER_c. The OPC for q > 1 blows up to infinity at ER_c, meaning that values of ER above ER_c should never be accepted, regardless of the decision time achievable and regardless of the SNR. The value of ER_c decreases as a function of q, indicated by the leftward shift of the maximum OPC value plotted as q

---

[1] Motivated by the COBRA (competition between reward and accuracy) theory of Maddox and Bohil (1998), Bogacz et al. (2006) also introduced another reward rate (Reward-Accuracy) function $RA(q) = RR - q\frac{ER}{D_{tot}}$. This does not fit the present data as well as RR_m, and we do not consider it further; however, it does fit the single-session data of Bogacz et al. (2010) better than RR_m (see Zacksenhouse et al., 2010).

increases in Fig. 1. For any SNR > 0, however, the pure DDM can achieve arbitrarily low error proportion at the expense of arbitrarily long decision times. Thus, it is only for SNR = 0 and $q > 1$ that the optimal strategy is to cease responding altogether.

Suboptimal performance in 2AFC: accounting for timing uncertainty

It is important to note that assuming participants can maximize RR_m entails the further assumption that they can estimate task parameters accurately and precisely. Naturally, however, they might instead be uncertain about task parameters. For instance, the temporal intervals that constitute a defining feature of many decision-making scenarios (e.g., response-to-stimulus intervals, response deadlines) are known to be represented with a level of endogenous uncertainty (Gibbon, 1977; Rakitin et al., 1998). Temporal information in itself is used adaptively in many temporal decision-making scenarios (e.g., Balci, Freestone, & Gallistel, 2009), but it may also be crucial for estimating other task-relevant information such as the reward rate (number of rewards per unit time). Thus, uncertainty in the representation of critical temporal intervals might also lead to "suboptimal" performance (and conservative decisions) by producing noisy estimates of reward rates.

In favor of this alternative, Bogacz et al. (2006) showed that the expected reward rate in our task is an asymmetrical function of threshold, which descends more steeply for thresholds below the optimum than for ones above it. Consequently, for a given level of uncertainty, it is better to overestimate the threshold than to underestimate it. This can explain suboptimal performance with slower responses (higher thresholds) without appeal to an inherent preference for accuracy. Although the asymmetry in reward rate as a function of threshold is most accentuated at higher signal-to-noise ratios, there may be other reasons to prefer higher than optimal thresholds, such as preventing the tendency to make anticipatory responses (cf. Simen et al., 2009). Zacksenhouse et al. (2010), furthermore, recently showed that an information gap decision strategy that maximizes the minimal reward rate achievable for a given level of uncertainty in time estimation fits the data presented in Bogacz et al. (2010) better than the optimal performance curve, RR_m. These observations motivated us to investigate the relation between timing uncertainty and the degree of deviation from optimal performance in the current study.

Suboptimal performance in 2AFC: failure to set independent thresholds

When the same participant is tested with multiple levels of SNR blocked into groups of trials with the same SNR,

setting optimal decision thresholds independently for each different SNR condition requires participants to decide which SNR is presented in a given block. Although the possible mechanisms by which humans estimate SNR is outside the scope of this paper, we note that it likely requires some sampling time to reach an accurate, if still noisy, estimate. One cognitively affordable strategy in the face of such cost and uncertainty is to use a single decision threshold and attempt to maximize the average reward rate across multiple SNRs. This can be a good strategy especially in contexts in which setting a single threshold results in only negligible reductions in the maximum possible expected reward rate. When the cognitive cost of encoding/retrieving/setting independent thresholds and/or sampling cost is considered, setting a single threshold might even turn out to be preferable to setting independent thresholds. In this paper, we therefore examined whether participants adopted single or multiple decision thresholds and evaluated them within the framework of optimality.

Methods

Participants 17 adults (6 males and 11 females), aged 18–30 years, were recruited via announcements posted online and around the Princeton University campus. Participants were paid 2 cents for each correct choice. The experiment comprised 13–15 daily, hour long sessions. Twelve participants completed 15 sessions, one participant completed 14 sessions, and four participants completed 13 sessions. The experiment was approved by the Institutional Review Panel for Human Subjects of Princeton University and all participants provided written consent for their participation.

Stimuli and apparatus The visual stimulus consisted of a field of randomly moving dots, each of which appeared within an aperture of 3 inch diameter in the center of the screen. Dots were white 2×2 pixel squares displayed against a black background. Stimuli were viewed from approximately 2 feet (c.61 cm) (approximately 8° visual angle). On each trial, a fraction of the dots moved in a single direction over time, corresponding to that trial's correct direction, while the remaining dots were randomly repositioned over time. The motion direction (rightward or leftward) was assigned randomly with equal probability. The display was generated in MATLAB on a Macintosh computer with a CRT monitor, using the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997) and DotsX software coded by Josh Gold and his laboratory. Responses were collected with a standard computer keyboard.

Procedure Each session consisted of blocks of decision-making trials, plus blocks of either a signal detection task or

an interval timing task, details of which are provided below. The first four sessions consisted of a 2-min block of practice of a free-response dot motion discrimination task with 50% motion coherence. This was followed by practice on a peak-interval timing task (10 training and 10 test trials, using the procedure of Rakitin et al., 1998). Participants did not receive monetary reward during practice. Participants then performed five 5-min blocks of the free-response dot motion discrimination task, each with a different level of motion coherence, followed by 10 training trials and 40 test trials of the peak-interval timing task. Participants could take a break of up to 5 min at the end of the peak-interval timing trials. Finally, participants underwent two 2-min blocks of a signal detection task. From the 9th or 13th session onwards (randomly assigned per participant), peak-interval and signal detection trials were replaced with two blocks of a response-deadlined dot motion discrimination task (data not described here), in which the participants earned scores only for accurate decisions that were made prior to the deadline.

1. Free response dot motion discrimination task: Participants were instructed to decide as quickly and accurately as possible on each trial whether the motion was leftward or rightward. Participants pressed the 'M' key on the computer keyboard when they decided the motion was rightward and the 'Z' key when the motion was leftward. Each correct response was signaled by a short audio tone whereas there was no feedback for errors. Participants were awarded a point (corresponding to 2 cents) for every correct response. There was no monetary or time penalty for errors. Trials were separated by a response-to-stimulus interval (RSI) sampled from an exponential distribution (with lower and upper bounds) with a mean of 1 s. Premature responses (RT < 100 ms, including anticipatory responses prior to stimulus onset) were followed by a buzzing sound and penalized by a 4-s timeout period. The cumulative score was displayed in the center of the screen after every five trials. Trials were blocked by coherence conditions. Five coherences of 0, 4, 8, 16 and 32% were used, with block order randomized across participants. Each block was 5 min long (median 187 trials, interquartile interval 31).

2. Signal detection task: In order to calibrate our estimates of the time taken by non-decision processes, we included a signal detection task. Participants were presented with a dot motion stimulus but were instructed to ignore the coherence and respond as soon as they saw the stimulus. For one of the blocks, they were instructed always to press the 'M' key whenever they saw any stimulus, and in the other, they were instructed always to press the 'Z' key. Participants were

rewarded with 2 cents for each correct key press that was not premature. Premature responses were penalized as described above.

3. Peak-interval timing task: In order to characterize participants' inherent timing precision, we included task blocks on a timing task previously used with humans (Rakitin et al., 1998). An interval of either 1, 2, 4 or 8 s was used in each session. The presentation order of the interval times was counterbalanced across participants.

Training phase The training phase comprised 10 trials. Participants were provided written and verbal instructions stating that a blue square would appear in the middle of the screen and change color to magenta after a certain period of time. They were asked to pay attention and learn this target interval. Participants were instructed not to count, or to use any process of subdivision such as foot-tapping. In order to signal the target interval on each training trial, a blue rectangle changed color to magenta after a target duration had elapsed since the onset of stimulus presentation. Target durations were held constant in each session. In order to discourage explicit counting, small digits appeared at random intervals on the center of the rectangle, and participants were instructed to read these digits out aloud. At the end of the trial, the screen went blank, signaling that a new trial could be started by a key press.

Testing phase The testing phase, which followed the training phase, comprised 40 trials. Test trials were similar in all respects to the training trials except that the blue rectangle presentation was longer than the target time interval and the rectangle did not change color during these trials. Participants were provided written and verbal instructions to produce responses centered around the target time interval that started with the onset of stimulus presentation (Rakitin et al., 1998). As in the training phase, they were instructed not to count, or to use any process of subdivision and to read the digits presented on the screen out aloud (see training phase). After each trial, a histogram was presented to the participant displaying the distribution of responses on the previous trial along with the target interval on a relative scale. Trials were automatically terminated if three times the interval period had elapsed.

Data analysis

DDM fits Individual participants' 2AFC data were first pooled across sessions 10 and greater and then were fit by the DDM using the diffusion model analysis toolbox (DMAT) in Matlab (Vandekerckhove & Tuerlinckx,

2008). Initially, we fit three different drift-diffusion models to the data. In fitting the pure DDM to the data, the decision threshold (z), drift (A), and non-decision times (T_0) were allowed to vary across conditions, while bias in the starting point and inter-trial variability in the core parameters were not permitted. In fitting two extended versions of the DDM, we allowed for inter-trial variability in the core parameters. In one of these model fits, no bias was allowed in the starting point (we refer to this model as the constrained extended DDM) whereas it was allowed in the other (we refer to this model as the full extended DDM). Pure DDM fits were also conducted for each individual participant's data and are presented wherever relevant.

Given its extra parameters, the extended DDM naturally fits data better than the pure DDM, but the superior fit quality of the more complex model holds even when its extra parameters are taken into account using the Akaike Information Criterion test for small samples (AICc) or the Bayesian Information Criterion test (BIC). Indeed, for the highest three coherences, the constrained extended DDM (AIC$_c$: 215,597.75, BIC: 215,758.20) outperformed the pure DDM (AIC$_c$: 216,789.31, BIC: 216,869.54) and the full extended DDM (AIC$_c$: 214,799.77, BIC: 214,986.97) outperformed the constrained extended DDM. Similar results held for the 0% and 4% coherence conditions (AICcs: 165,461.44, 165,346.06, and 165,081.40 and BICs: 165,512.35, 165,447.89, and 165,200.20 for the pure, constrained extended, and the full extended DDM, respectively).

In a second set of analyses by which we examined the plausibility and optimality of a single decision threshold setting across coherences, we fit the same models without allowing the decision threshold to vary across coherence conditions. As before, we fit only the highest three coherences in order to minimize the effects of fast-guesses and other contaminants that are clearly present in the 0 and 4% coherence conditions (see Results section for discussion).

Nevertheless, fitting procedures are typically noisy, and there is reason to believe that noisy parameter estimates can obscure the degree to which data in fact approximate optimality (see Figs. 4 and F1 in Simen et al., 2009). Furthermore, using the extended DDM complicates the process of generating optimality predictions, since with this model these predictions can only be achieved through numerical simulations. The parameter-free optimal performance curve that is derived from the pure DDM, in contrast, can be compared to data without any fitting at all. In this paper, we primarily used the parameter-free approach (i.e. OPC) but also validate it by a model fitting approach. We also conducted optimality analysis for the constrained and full extended DDM by numerical simulations and present it in the supplemental material.

Peak responding Given that the mean response-to-stimulus interval was 1 s in the free-response dot motion discrimination experiment, we used the 1-s peak interval data to estimate participants' uncertainty about the temporal structure of the task. Median key press times for each trial were determined for 80 testing trials. The standard deviation of median response times was then divided by their mean in order to determine the coefficient of variation, a measure of timing precision, for each participant.

For multiple regression analyses, we used the data and q value estimates from sessions 2–13. The effects of session number and coherence level on accuracy and deviation from the optimal performance curve were tested using two-way repeated measures ANOVA. We used the Greenhouse–Geisser correction when the sphericity assumption was violated. An alpha level of .05 was used for all statistical tests. Bonferroni correction was used to correct for multiple comparisons (unless stated otherwise).

Results

We first examine the relation between observed decision-making performance and optimal performance as defined by the DDM, focusing on how this relation evolved over the course of training. We show that when performance was characterized in terms of decision times and accuracies (either separately for, or irrespective of, different coherences), the behavior of highly practiced participants conformed to patterns predicted by optimality for all but the lowest motion coherences. With model fits, we then show that for decisions that rely on sensory evidence participants on average adopted a nearly optimal single decision threshold across all signal qualities. We demonstrate that by setting a single threshold participants still closely approximated the maximum possible expected reward rate predicted by independent optimal thresholds. We further examine the alternative cost function RR_m (Eq. 4), in order to explain pre-asymptotic performance. Finally, we examine individual differences in interval timing ability as an explanation of residual deviations from optimality across all sessions.

Before presenting results, we note that response times in the 0 and 4% coherence conditions exhibited a different pattern than in higher coherences. Visual inspection of the response time distributions averaged across participants revealed bimodality particularly for these coherences (Fig. 1 of supplemental material). Bimodality was also present in around half the participants' individual data. The shorter mode of the response time distribution was well aligned with the mode of the signal detection time (SDT) distribution. Furthermore, the density of the short response

time distribution diminished with increasing signal quality. These observations suggest the presence of a mixture of integrative and non-integrative decisions, in which motion direction evidence is or is not accumulated, respectively. This is true particularly in the 0 and 4% coherence conditions (cf. Simen et al., 2009), where a threshold of nearly 0 would in any case be optimal (at 0% coherence, the optimal threshold is exactly 0). The greater distance between the 10th and 90th percentiles of the response time distribution for the 0 and 4% coherence conditions (compare Fig. 4, left and right panels) also reflects this nature of the response time distributions. Thus, the optimality analyses for DDM fits to the data collected from 0 and 4% conditions should be interpreted with caution. For this reason, and in order to avoid the contamination of DDM fits to higher coherences for which decisions were primarily integrative (i.e. 8–32%), we excluded 0 and 4% coherence conditions when fitting DDM with a single threshold. Other arguments regarding fitting the DDM are discussed in the supplemental material.

Comparison of empirical and optimal performance via OPC

Figure 2a shows average decision times as a function of average error proportions separately for different coherences and during four different periods in training (sessions 1, 2–5, 6–9, 10–13). Visual inspection of Fig. 2a suggests that performance approached the optimal performance curve over the course of training and reached at its asymptotic level after around five sessions. At steady state (i.e. sessions 6 and up), participants performed nearly optimally for the highest three coherences (i.e. 8, 16, and 32%) but suboptimally for the lower two (i.e. 0 and 4%). In order to assess the "steady state" performance, we compared the empirical decision times gathered from sessions 10–13 to the optimal decision times prescribed by the pure DDM for the corresponding mean empirical error proportions. Results corroborated our observations based on Fig. 2a. Comparison of the empirical to optimal decision times (computed for the mean empirical accuracies) separately for different coherences did not reveal any significant differences for the highest three coherences (all ps > .42, uncorrected), but did achieve significance for the 0% [$t(16)$ = 4.78, p < .01, corrected] and 4% [$t(16)$ = 3.22, p < .05, corrected] coherences. Null findings held after comparing the empirical decision times to the optimal calculated for accuracies one standard error of the mean (SEM) below and above the mean (Fig. 2a, horizontal error bars).

Figure 2b, c depicts an alternative characterization of the performance, namely the average decision times as a function of the error proportions irrespective of coherences (as presented and analyzed in the earlier theoretical work of Bogacz et al., 2006). Blocks from different training periods

were first sorted by accuracy in five categories (ERs = 0–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4, 0.4–0.5) irrespective of coherences. Decision times from these blocks were averaged for each participant separately for each accuracy bin and training period. These averages were used for summary statistics for corresponding error proportion bins. In line with Fig. 2a, visual inspection of Fig. 2b suggests the overall convergence of these empirical decision times on the optimal performance curve. Despite this overall improvement in performance (and in parallel to Fig. 2a) by the end of the training, decision times with high error proportions were still greater than optimal. In order to assess the "steady state" performance (now characterized irrespective of coherences), we compared the empirical decision times gathered from sessions 10–13 (averaged for each participant) to the optimal decision times prescribed by the pure DDM for the mean accuracy of each bin. These comparisons showed that for ER > 0.3, empirical decision times were significantly larger than the optimal decision times (both ps < .01, corrected).

In order to control for the contribution of the most difficult conditions in assessing these deviations, we also plotted data that excluded the 0 and 4% coherence conditions (Fig. 2c). Note that even for high coherences, several participants performed at accuracy levels between 0.3 and 0.5. After the exclusion of these two most difficult conditions, empirical data approximated the optimal performance curve much more closely and the empirical decision times were not significantly different from the optimal decision times for any bin (all ps > .06, corrected). Briefly, when the no-signal (0%) and difficult (4%) blocks were excluded, averaged performance converged on the optimal performance curve by the end of the experiment. In the supplemental material, we present a more detailed analysis of the correspondence between empirical and optimal decision times (see the orthogonal regression section).

In order to better characterize the improvement in performance over the course of training, we examined how deviations from the optimal performance curve changed from session to session separately for different coherence levels (thus paralleling the depiction in Fig. 2a). Here, deviations refer to the difference between the empirical decision time and the optimal decision time prescribed by the optimal performance curve for the corresponding ER (i.e. the vertical distance between the empirical and optimal decision times in Fig. 2). Figure 3a shows the cross-participant averaged acquisition curve separately for the different coherences for up to 13 sessions. In line with our previous analysis, visual inspection of Fig. 3a shows that decision times converged towards the optimal value for all coherences and reached the optimal performance curve for all but the two lowest motion
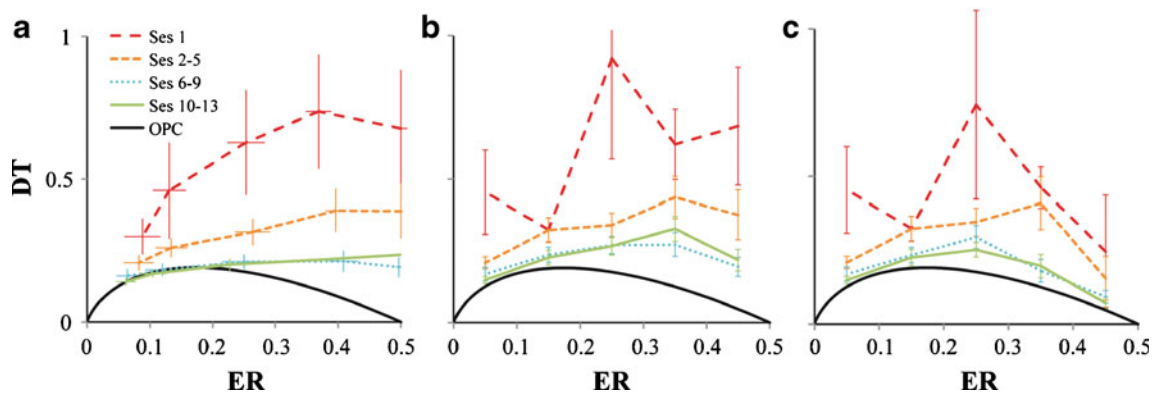
Fig. 2 a Mean empirical decision times grouped by coherence and shown separately for sessions 1, 2–5, 6–9, and 10–13. *Vertical error bars* SEM for decision times, *horizontal error bars* SEM for error proportions. The *inverted U-shaped bold curve* is the optimal performance curve of the pure DDM, (Eq. 3). **b** Average empirical decision times separated in five error proportion bins. Blocks were sorted by accuracy in five categories: ERs = 0–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4, and 0.4–0.5 irrespective of coherences. Decision times from these blocks were averaged for each participant separately for each accuracy bin and training period. These averages were used for summary statistics for the corresponding error proportion bins. **c** Same as in Fig. 2b but excluding 0 and 4% coherence conditions

coherences. Deviations from the optimal value remained higher for the lowest two coherences throughout the training. Results of repeated measures two-way ANOVA analysis corroborated these observations. Coherence and training sessions were found to have a significant effect on the deviation from optimality, $F(1.46, 23.41) = 17.17$, $p < .001$ and $F(1.80, 28.79) = 4.75$, $p < .05$, respectively. There was no significant interaction of coherence and training sessions, $F(3.96, 63.36) = 1.88$.

Figures 2b, c, and 3a depict only the change in distance from the optimal performance curve, a measure that is blind to possible improvements in accuracy throughout training: namely, horizontal leftward shifts along the ER axis. Figure 2a on the other hand suggests small improvements in accuracy over the course of training particularly for the higher coherences. To better address this possibility, Fig. 3b shows the session-by-session, cross-participant averaged accuracies, plotted separately for different coherence conditions. The plots reveal that accuracies remain stable (approximately constant) throughout training, although they differ significantly across coherences, being higher for higher coherences. Results of repeated measures two-way ANOVA analysis (excluding the 0% coherence condition) corroborated these observations. Coherence had a significant effect on accuracy, $F(2.06, 32.88) = 158.22$, $p < .0001$ while training sessions had no significant effect on accuracy[2], $F(3.10, 49.52) = 1.35$. There was no significant interaction between coherence and sessions, $F(7.56, 120.99) = $
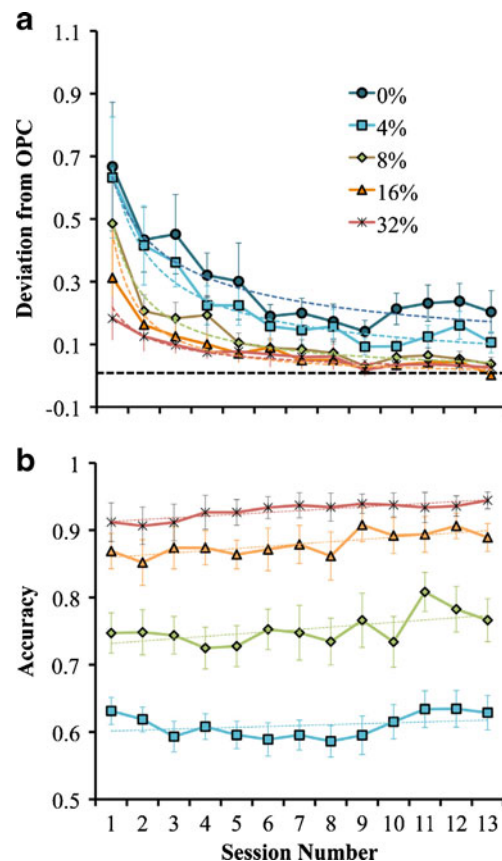

Fig. 3 Session by session cross-participant averaged **a** signed deviations from the optimal performance curve in each coherence condition, **b** accuracies for different coherence conditions. *Dashed curves* in (**a**) are power function fits to the averaged data. *Dashed lines* in (**b**) are linear regression fits to the average accuracy data. Data are depicted up to the 13th session

[2] Figure 3b, however, suggests a slight increase in accuracy (see also Fig. 2a) for some coherences (i.e. 8 and 16%). In contrast to all others, one participant performed at chance for all the coherence levels for the first 4 sessions, and these slight, non-significant enhancements in accuracy were primarily due to this participant. Training sessions did not have an effect on accuracy, even when each motion coherence condition was analyzed separately (all $ps > .15$).

1.15. Figure 3 of supplemental material presents plots that are analogous to Figs. 2b and 3a but depict individual participant data that corroborates this finding. The data therefore suggest that reductions in response time occurred over training without concomitant reductions in accuracy (cf. Dye, Green, & Bavelier, 2009).

Comparison of empirical and optimal performance via DDM fits

Deviations from the optimal performance curve are indirect measures of the degree to which model parameters approximate their optimal values. In contrast, fitting model parameters allows direct comparison to theoretically optimal values (although the comparison then depends on the quality of the fitting procedure, which is not the case for the parameter-free OPC analysis). In order to depict the quality of pure DDM, extended DDM, and constrained extended DDM fits (see Data Analysis), Fig. 4 shows quantile-probability plots (Ratcliff, 2001) separately for 0–4% (Fig. 4a) and 8–32% (Fig. 4b) coherences. Model comparison statistics for pure DDM, extended DDM, and constrained extended DDM are reported in the Methods section.

Since the pure DDM fits the data fairly well and since the only known, closed-form expression for an OPC is derived from the pure DDM (Eq. 3), we will conduct optimality analyses primarily based on the parameter estimates from pure DDM fits. In the supplemental

material, we present simulation-based optimality analysis for the constrained and full extended DDM. We first present our statistical analysis of the parameter values gathered from pure DDM fits. Pair-wise comparisons of drift rates estimated from individual participants' data across different coherences (Fig. 5a) revealed significant differences across all coherences (all $ps < .01$, corrected).

We then calculated the optimal threshold given the drift rate separately for different coherence levels. This computation was done in two different ways to ensure that our results were not an artifact of any particular averaging method: (1) separate thresholds were computed for a set of individual drift rates, each estimated from an individual participant's data, and these thresholds were then averaged across participants (Fig. 5b, Definition I, solid red line); and (2) a threshold was computed for the across-participant average of the individual drift values (Fig. 5b, Definition II, dashed orange line). Figure 5b shows that the estimated thresholds were far from the optimal threshold for the 0 and 4% coherence conditions; in contrast, they were clustered around the optimal threshold for the higher coherence conditions. In Fig. 5c, we show averaged fitted thresholds separately for highest three coherence levels. On the same figure, we also show the mean optimal threshold (Definition I, calculated for each participant, solid red line) along with optimal thresholds calculated for the mean drift rate (Definition II, dashed orange line). For this plot, we excluded those participants with thresholds that were two standard
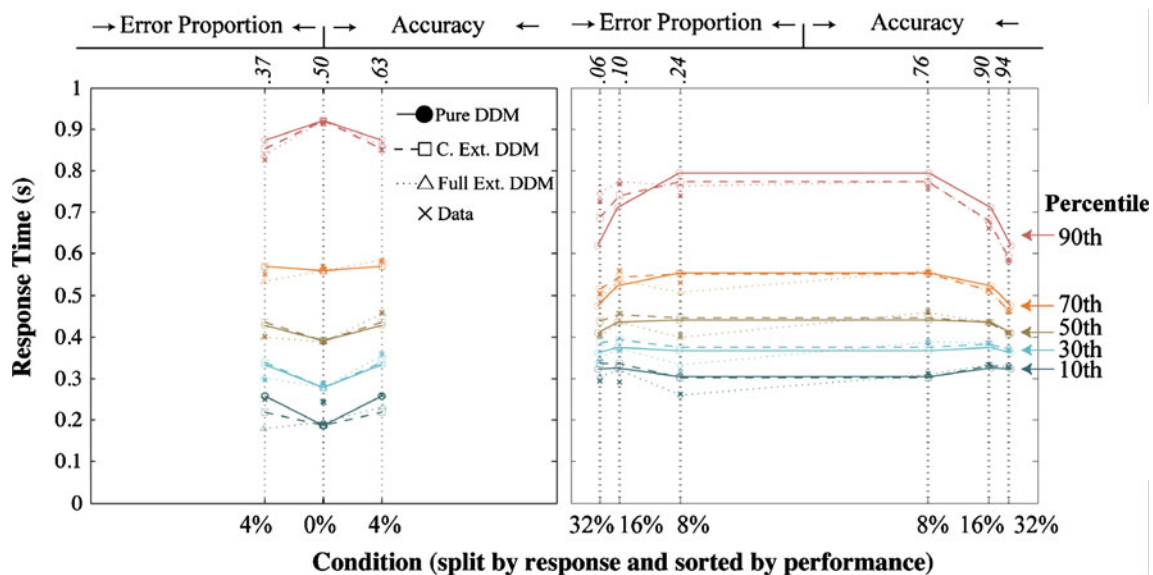

Fig. 4 Quantile probability plots for all five coherences at the 10th, 30th, 50th, 70th, and 90th percentiles of response times separately for 0–4% (*left panel*) and 8–32% (*right panel*) coherences. *Points* depict the empirical data. *Solid lines with circles* depict pure DDM fits, *dashed lines with squares* depict constrained extended (*c. ext.*) DDM fits, and *dotted lines with triangles* depict full extended DDM fits. *Vertical dashed lines* indicate different coherences, for errors (*left*) and

correct responses (*right*). Error proportions and accuracies corresponding to different coherences are presented at the top. The full extended DDM fits the data best followed by the constrained extended DDM. The pure DDM predicts equal mean error and correct response times, which also provides a fairly good, although perfectly symmetrical, fit to the data
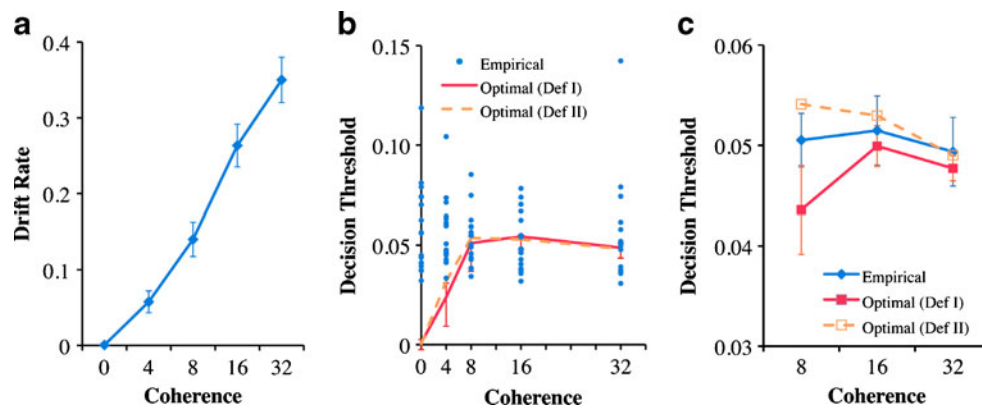
Fig. 5 a Average drift rates estimated separately for individual participants' data for different coherences. **b** Empirical decision thresholds plotted along with median optimal thresholds calculated for drift rates estimated from individual participants' data (*Def I*, see text) and optimal thresholds calculated for average drift rate estimates (*Def II*, see text). Each *blue point* corresponds to an individual participant's data in each of five different coherence levels. *Error bars* IQI **c** Average thresholds (after excluding two outliers) along with the average optimal thresholds (*Def I*, for the corresponding participants) and optimal thresholds estimated from the averaged drift rate (*Def II*, for the corresponding participants) separately for the highest three coherences. Refer to text for these definitions. *Error bars* SEM

deviations above and below the mean separately for each coherence (thus excluding one participant from the 8% and one participant from the 32% coherence condition). When these threshold estimates were compared to the two different optimal threshold curves (Definitions I and II) separately for different coherences, there were no significant differences between the empirical and optimal values for the 8, 16, and 32% coherences (all $ps > .84$, corrected), whereas they were reliably different for the 0 and 4% coherences (all $ps < .01$, corrected). Briefly, our comparison of the empirical and optimal performance via decision thresholds estimated from the fits corroborated our comparison via the OPC.

Despite both empirical and optimal thresholds qualitatively following the same curvilinear pattern across the highest three coherences, there were no statistically significant overall differences (one-way between-subjects ANOVA) among either the optimal thresholds $F(2, 46) = 1.26$, $p = .29$ or the empirical thresholds $F(2, 46) = .11$, $p = .90$ across these coherences. Thus, it seems quite possible that participants adopted a single threshold rather than multiple thresholds for these coherences. This is particularly likely given that the optimal thresholds themselves were not significantly different from each other across the top three coherence conditions. Furthermore, the relationship between decision time and error proportion observed in Fig. 2a qualitatively suggests a single threshold across all coherences, including the lowest two, for which a large drop in both estimated thresholds and empirical decision times was expected but was observed only for a subset of participants in a subset of the trials.

DDM fits with single thresholds

Based on these observations, we also fit individual participants' data without allowing the decision threshold

to vary across coherences. In order to minimize the effect of contaminants, we fit only the data from the higher three coherences. In all cases, based on AIC$_c$ and BIC values, models with multiple thresholds fit the data equally well or better than models with single thresholds. Model comparison statistics for different training periods are presented in Table 1 of the supplemental material. In order to compare the single versus multiple threshold models for these conditions, we further conducted a split-half cross-validation analysis. For this analysis, we first pooled the data across participants for sessions 10 and above and randomly sampled half the data points (without replacement). The DDM with single and multiple thresholds was then fit to the first, randomly sampled half of the pooled dataset. The best fitting parameters of the two models were then used to evaluate AIC$_c$ and BIC values for the other half of the data. This procedure was conducted 100 times, and the mean and median difference between the AIC$_c$ and BIC values usually favored the single threshold model over the multiple threshold model (mean ΔAIC$_c$ = 3.00, mean ΔBIC = 7.88 favoring the single threshold model). For the highest three coherences at least, these results reinforced the notion that a single threshold was used by most participants and optimized for the ensemble of task conditions, rather than that a separate threshold value was encoded, retrieved and finely tuned for each coherence condition. Subsequent results will therefore be based on single-threshold estimates based on fits to the highest three coherences. Later, we also discuss the possibility that some participants transition into a distinct state of non-integrative responding (with starting point-to-threshold distances equal to 0) at very low coherences.

We compared the best fitting single thresholds with optimal single thresholds (computed numerically) at steady

state (i.e. Sessions 10 and above). Fitted single thresholds were not significantly different from the optimal single thresholds, $t(16) = 1.71$, $p = .11$. As stated earlier, the particular relation between decision times and error proportions that is revealed when data is categorized based on coherences (Fig. 2a) qualitatively suggested single decision threshold setting. Thus, we also fit the empirical decision times and corresponding error proportions with a single threshold given the average drift rate estimates from DDM fits with single thresholds (see Fig. 6).

We fit data collected from four different training periods twice: once based on the highest three coherences (solid lines) and once based on all coherences (dotted lines). In these fits, we used drift rates averaged across participants separately for different coherences. Drift rates for the highest three coherences were gathered from fits with the threshold constrained to be equal across conditions. Drift values for the lower two coherences (i.e. 0 and 4%) were estimated from fits allowing thresholds to vary across conditions. Figure 6 shows that despite systematic discrepancies in the estimation of observed error proportions and a bad fit to the data from session 1, a single, cross-coherence threshold per training period could account for the observed relation during the majority of the training. Fitting both decision threshold and drift rate rather than using the drift rate estimates gathered from the DDM fits revealed very similar fits to the data. We note also that a better fit to these average DT/average ER data points can be obtained with lower drift and threshold estimates than were recovered by model fits to the complete response time distributions,
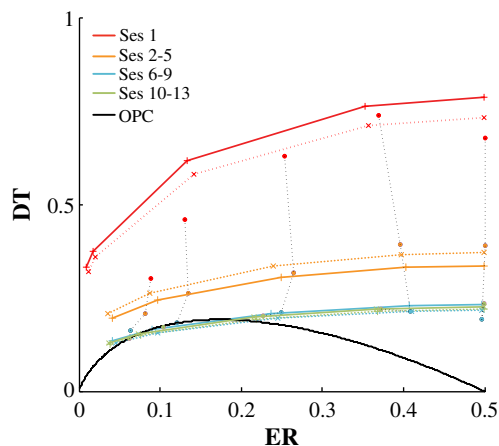

Fig. 6 Single-threshold fits in speed-inaccuracy space (the same data is presented in Fig. 2a). *Solid* and *dotted horizontal lines* are fits with a single threshold to the empirical decision time and error proportion data (*filled circles*). *Dotted black lines* connect the empirical data points from different sessions that come from the same coherence condition. Fits depicted by *solid lines* were computed based on drift rates solely for the highest three coherences. Fits depicted by *dotted lines* are based on drift rates for all five coherences. Note that the fits to the data from the last two training periods (sessions 6–9 and 10–13) superimpose

suggesting some degree of parameter inflation in the DDM fits (cf. similar evidence of parameter-estimation bias during DDM-fitting in Simen et al., 2009). The leftmost points of the solid lines under these parameterizations shift leftward as the session number increases, as do the empirical data.

Comparison via proportion of maximum expected reward rate

So far, we have characterized performance directly and indirectly based on the correspondence between optimal and estimated DDM parameters. An alternative way to characterize performance within the framework of optimality is to examine the proportion of the maximum possible expected reward rate secured by the participants. After all, what participants should have valued most in the task (if instructions were followed) was the monetary reward they earned.

The proportion of maximum possible expected reward rate was computed both liberally and conservatively. For the liberal computation, we simply divided the expected reward rate given the fitted threshold by the maximum possible expected reward rate for that coherence and participant. Note that this approach assumes that the worst thing a participant could do in the task is to set the decision threshold to infinity and thus not respond (thus a reward rate of 0). On the other hand, participants could simply choose not to integrate evidence and to respond as soon as they detect a stimulus (i.e., to respond *non-integratively*), which would result in reward rate that would be expected with a decision threshold of 0. In order to prevent this possibility from skewing our assessment of how closely participants approached optimal performance, we also computed the proportion of maximum earnings in a conservative manner: namely, we computed the proportion of the difference between the expected reward rate at the fitted threshold and the expected reward rate at a threshold of 0 to the difference between the maximum possible reward rate and the reward rate expected at a threshold of 0, $(RR_{emp} - RR_{z=0})/(RR_{max} - RR_{z=0})$. We computed these values for the best fitting single threshold as well as for the multiple-threshold fits.

Regardless of the type of analysis (conservative or liberal) and the type of fit (single or multiple thresholds) participants secured more than 95% (mean 98%) of the maximum possible expected reward rate for the highest three coherences (taking independent optimal thresholds for each coherence as the reference). This value on average (evaluated only for multiple-threshold fits) decreased to 90 and 78% for the 4 and 0% coherences, respectively. When performance was quantified based on the reward rate averaged across the highest three coherences (dashed red

curves in Fig. 7), participants overall secured 98% of the maximum possible expected reward rate (see Table 2 of supplemental material). This value decreased to 94%, when 0 and 4% conditions were also included in the computation of the average reward rate.

Table 2 of supplemental material presents these proportions also for multiple thresholds. Briefly, participants were nearly optimal when performance was characterized in terms of the proportion of reward rate. Importantly, the use of a single threshold would have been only negligibly costly to participants. Table 2 of the supplemental material summarizes the output of this analysis for sessions 10 and above. The proportion of maximum expected reward rate earned, based on parameter estimates from the pooled data, were very close to the proportions derived from fits to individual participants' data.

Evolution of critical DDM parameters over the course of training

To get a clearer picture of how drift rate and decision threshold estimates evolved over the course of training (cf. Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009), we fit the pure DDM to the session-by-session pooled data separately for each coherence (Fig. 8). In line with the average acquisition curves for deviations from the optimal performance curve, decision threshold estimates decreased over the course of training (Fig. 8, left panel). There were also local qualitative resemblances between these two plots (compare with Fig. 3a). In contrast with decision threshold, drift rates increased over the course of training, indicating improvement in signal-to-noise ratio, particularly for the high coherence conditions (Fig. 8, right panel)[3]. This is notable because it implies that the observed improvements in participants' earnings derived from an increase in speed without a concomitant change in accuracy, despite the evidence of increasing SNR over training.

Accounting for deviations from optimality: favoring accuracy over reward rate

One account for suboptimal performance in our task could be an emphasis that participants place on accuracy over reward rate, and the convergence of decision times onto the

optimal performance curve with practice could be interpreted as a decrease in this emphasis on accuracy over time. To address this possibility, we examined the session by session changes in q value estimates from our RR_m fits (see Fig. 9a and Fig. 3c of supplemental material); we also show concomitant increase in rewards. In parallel to our previous observations of diminishing deviations from the optimal performance curve and optimal decision thresholds with practice, q value estimates diminished over the course of training. Corroborating this observation, repeated measures one-way ANOVA revealed a significant effect of sessions on the q value, $F(4.44, 70.98) = 4.81$, $p < .01$.

Figure 9b shows the median q parameter fits based on the best fitting single threshold to the highest three coherences as well as the median normalized deviation from the optimum single threshold [i.e. (empirical–optimal)/optimal]. Consistent with Fig. 9a, b shows that q values and normalized deviations from optimal single thresholds decreased together over the course of training to negligible levels. In fact, at steady state, the q values and normalized deviations from the optimal threshold were not significantly different from 0; $t(16) = 1.36$, $p = .19$ and $t(16) = 1.40$, $p = .18$, respectively.

For each individual participant, Fig. 10 shows the reward rate curve averaged across the highest three coherences (dotted blue curves) and the optimal decision threshold (vertical dashed blue line) for the corresponding participant. On the same plots, we show the best fitting single threshold to those coherences (vertical solid red line) and the modified reward rate (RR_m) curve for which this observed single threshold was the optimum single threshold (solid red curves; see Fig. 9b for depiction of derived values). Consistent with our statistical analysis based on q values and normalized deviations from the optimal threshold, Fig. 10 suggests that, except in a few cases late in training, participants performed very close to optimally assuming that they selected a single threshold. Note that both vertical lines terminate at points on the blue average reward rate curve with very close reward rate values.

Accounting for deviations from optimality: temporal uncertainty

Rather than deriving from an alternative cost function such as RR_m, suboptimal performance could instead reflect the effects of timing uncertainty (Zacksenhouse et al., 2010). As discussed above, such timing imprecision might in turn introduce variability and bias in threshold estimation. Given the asymmetry of the reward rate curve around the optimal threshold (e.g., Figs. 7 and 10), it is optimal to distribute threshold estimates around a mean that is higher than the optimal value. In order to determine whether there was any effect of temporal uncertainty on decision-making

---

[3] In both Figs.3a and 8, there is an evident change in the performance of the participants at around sessions 10–12. This is possibly due to the indirect effects of the inclusion of response-deadlined dot motion discrimination blocks at session 9 and thereafter (data from which are not analyzed here). Nevertheless, the comparison of the deviation from the optimal performance curve and fitted q values from sessions 10–12 separately from the preceding sessions (starting from session 7), did not reveal any significant differences (all $ps > .05$).
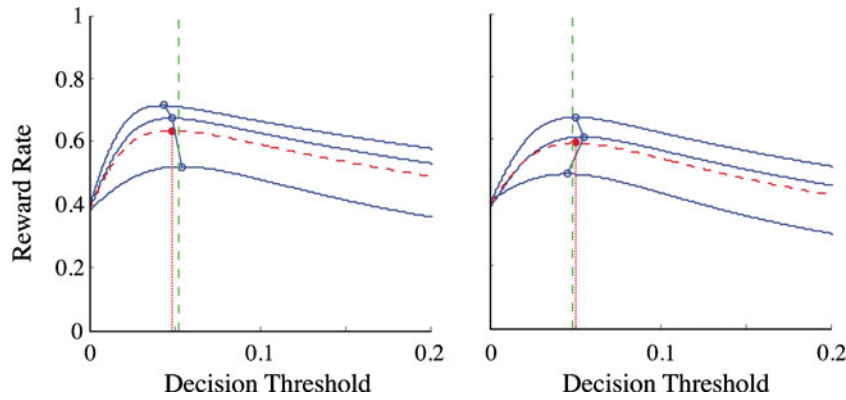
Fig. 7 Each plot corresponds to an individual participant's data. These plots show the reward rate curves as a function of decision thresholds computed from fits to the highest three coherences (i.e., 8, 16, and 32%) without allowing the decision threshold to vary. *Solid blue curves* are the pure DDM's expected reward rate curves for fitted drift and T_0 estimates, with the drift value fitted separately for each coherence level. The *dashed red curve* is the average of the three reward rate curves for the high coherences. The *solid line with empty circles* is the envelope of the optima that connects the peaks of each curve, with each circle denoting the optimal decision threshold and its expected reward rate for the corresponding motion coherence condition. The *red filled circle* and the *vertical dotted red line* denote the optimum single threshold. The *green vertical dashed* is the best fitting single threshold to highest three coherences

performance, we therefore conducted two sets of hierarchical regression analyses, assessing the degree to which q value and timing uncertainty predicted the suboptimal performance.

In the first, we adopted a conservative approach and first entered the q estimates from the RR_m model (estimated directly from the empirical deviation from the optimal performance curve) as the primary predictor of the vertical deviations from the optimal performance curve, and then added the timing coefficient of variation (CV) and signal detection time in the model (stepwise) in the second block. The hierarchical regression revealed a good fit ($R^2 = .88$). ANOVA revealed that the overall model (q and CV) was significant, $F(2, 14) = 48.85$, $p < .0001$. Adding the CV accounted for additional variance, $\Delta R^2 = .07$. Both accuracy weight q ($\beta = .75$, $p < .0001$) and CV ($\beta = .3$, $p < .05$) were influential predictors, whereas signal detection time ($\beta = .05$) was not. Figure 11 shows the multiple regression fit to the data. In a follow-up regression analysis, we first entered CV as the primary predictor of the deviation from the optimal performance curve and then entered q estimates in the model. This analysis revealed an R^2 of .45 for CV itself, $F(1, 15) = 12.1$, $p < .01$. Adding q accounted for additional variance, $\Delta R^2 = .43$. These results suggest that CV and q shared considerable variance. We did not find any reliable relation between the deviations from optimality and the uncertainty about the remaining three peak-interval durations (2, 4, and 8 s), which were all longer than the mean response-to-stimulus interval used in the decision-making task.

The same results held when this analysis was conducted using temporal uncertainty and q value estimates from single thresholds fits (Fig. 10) as predictors of the normalized deviation from the single optimum threshold. The details of this analysis are presented in the supplemental material.
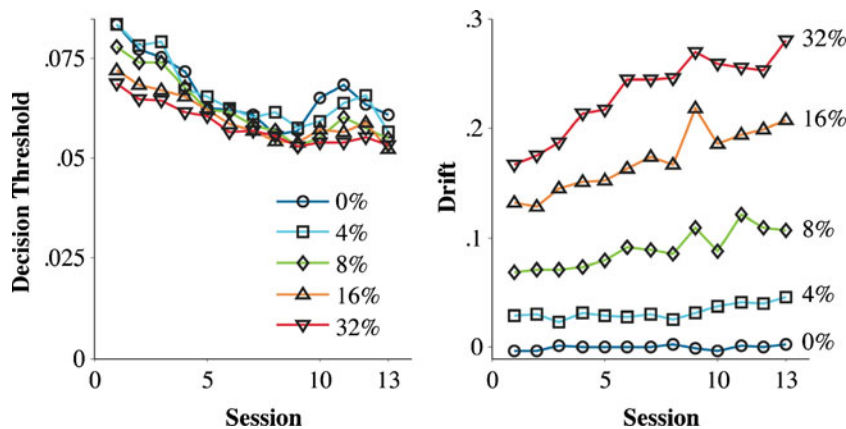

Fig. 8 Session-by-session decision threshold (*left panel*) and drift rate (*right panel*) estimates gathered from the pure DDM fits to the data pooled across participants

Fig. 9 **a** *q* values from the RR$_m$ fits (*solid line*) along with the session-total monetary rewards (*dashed line*). Data are depicted up to the 13th session. **b** *q* values and normalized deviation from the optimal single decision threshold for four different training periods (sessions 1, 2–5, 6–9, 10 and above). Values reported in (**b**) were computed from DDM fits with a single threshold to the highest three coherences
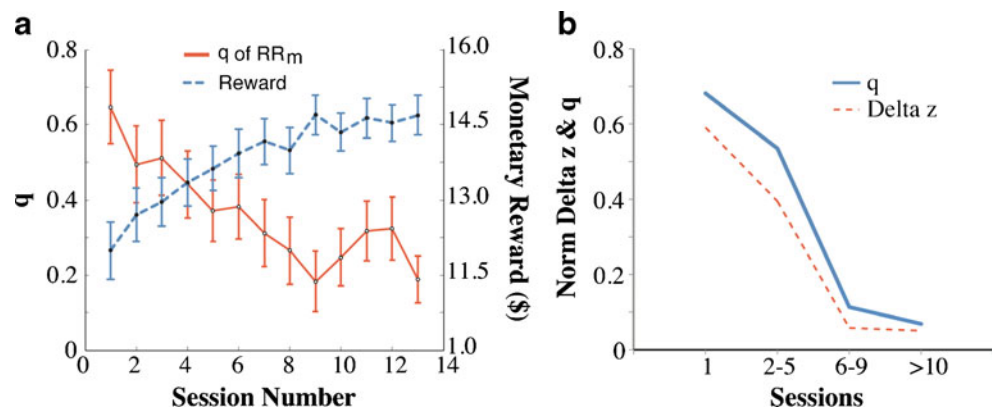
Discussion

In many situations, success depends on speedy simple decisions. However, despite their intrinsic simplicity, decision makers typically have limited access to information, so that time pressures imposed by the environment compromise their accuracy. The drift-diffusion model predicts optimal speed–accuracy trade-offs in 2AFC tasks administered with fixed signal-to-noise ratios and response-to-stimulus intervals in each block of trials (Bogacz et al., 2006). In a recent study of this prediction, 70% of human

participants were found to adopt suboptimal speed–accuracy trade-offs in single session two-alternative forced-choice tasks (Bogacz et al., 2010), while in a second experiment, extended practice with relatively high signal-to-noise ratios led to nearly optimal performance (Simen et al., 2009; see also Starns & Ratcliff, 2010). We expanded these investigations by further lengthening the training (up to 15 sessions) and testing a wide range of signal-to-noise ratios (0–32% coherence). We also conducted a new investigation of the relationship between the deviation from the optimal trade-off and uncertainty in representing the temporal
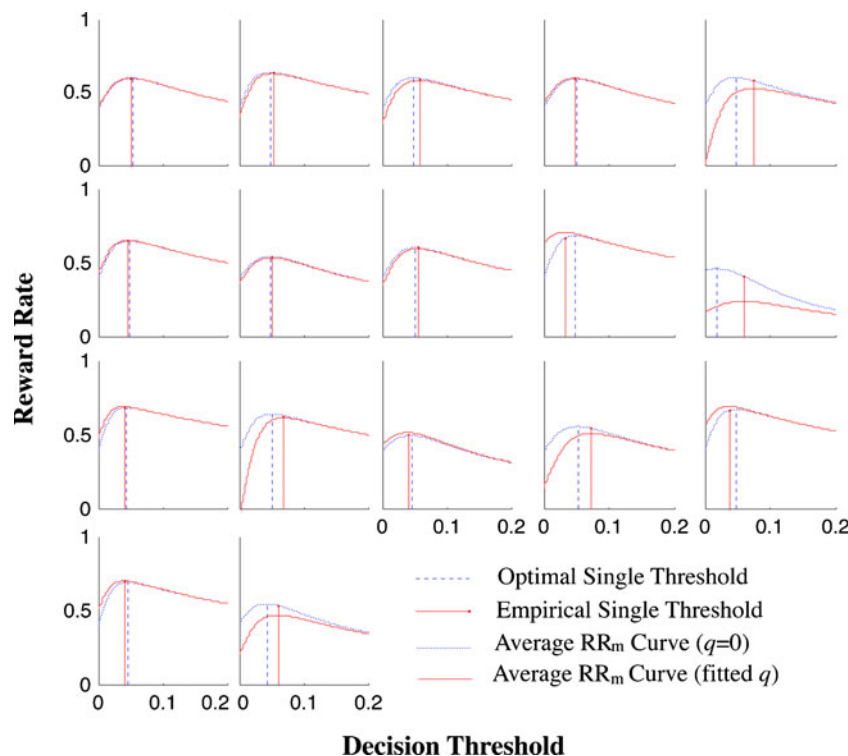

Fig. 10 Reward rate curves computed from fitted DDM parameters to the highest three coherences during sessions 10 and above. For the fits presented here, the decision threshold was constrained to be equal across coherences. Each plot corresponds to a different participant. *Dotted blue curves* are the reward rate curves estimated from the best fitting DDM parameters (*q* = 0) along with the optimum single threshold denoted by the *dashed blue vertical line*. The *vertical red solid line* shows the best fitting single threshold to highest three coherences. *Solid red curves* are the modified reward rate (RR$_m$) curves for which the best fitting single decision threshold was the optimal decision threshold under the assumption that participants were optimizing an alternative reward rate function (Eq. 4)
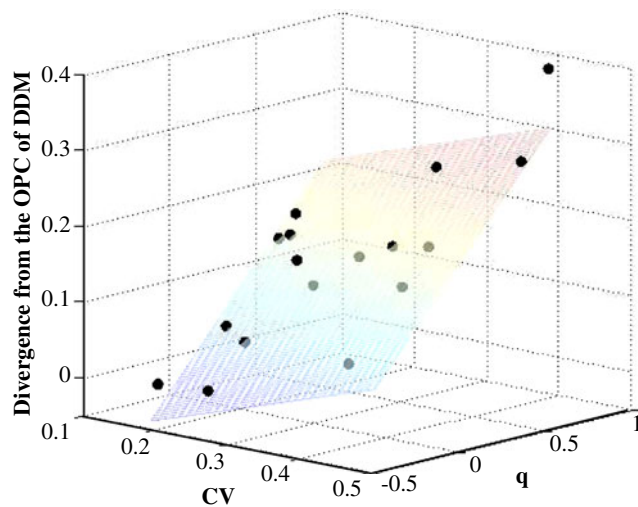
Fig. 11 Deviations from OPC of individual participants averaged across all coherences, plotted as a function of coefficient of variation and accuracy weight q of the RR_m. The fitted multiple regression plane is also plotted

structure of the task (i.e. mean response-to-stimulus interval).

Participants in our study converged on the optimal speed–accuracy trade-off over the course of training, without significant increases in accuracy. This was a result of an interaction between decreasing decision thresholds and increasing drift rates (perceptual learning) over the course of training (see also Dutilh et al., 2009). In other words, participants did not increase the reward rate simply by speeding their decisions but also by improving their perceptual abilities over the course of training in such a way that decreased response times traded off against improvements in accuracy.

Irrespective of how we characterized performance, either as a function of coherence or error proportion, in the last 8/15 sessions observed speed–accuracy trade-offs were close to optimal in the case of higher signal qualities (i.e. coherence ≥8%), whereas participants still appeared to set their decision thresholds well above the optimal threshold in the difficult (i.e. 4%) and no-signal conditions (i.e. 0%). When the performance was categorized solely as a function of error proportions and low coherence data were excluded from the analysis (see also Bogacz et al., 2006), optimal performance was observed over all error proportions. In summary, following sufficient training, humans adopted near-optimal speed–accuracy trade-offs for a wide range of signal qualities whereas deviations from optimality persisted for the difficult and no-signal conditions.

Interestingly, the relation revealed by categorizing performance by coherence (Fig. 2a) suggested the use of a single threshold across coherences rather than independent decision thresholds for each coherence. Drift diffusion model fits with varied decision thresholds further supported

this conclusion as neither empirical nor optimal decision thresholds were found to vary across the highest three coherences. Importantly, we found that adopting a single threshold rather than multiple thresholds resulted in negligible decrements in the maximum possible expected reward rate. We further showed that, under both single and multiple threshold assumptions, participants secured around 98% of the maximum possible expected reward rate (taking multiple thresholds as our reference for optimality). Briefly, participants adopted nearly optimal single, or possibly multiple, decision thresholds such that they secured almost all of the maximum possible expected reward rate.

Adopting optimal, independent decision thresholds for different coherence conditions would appear, in any case, to require the ability to estimate signal qualities accurately from a limited number of trials. This is a computationally difficult learning problem, however. Furthermore, exercising such an ability might reasonably be expected to impose costs in terms of effort and decision speed. When such costs are taken into account, and given the negligible effect of precise, condition-specific threshold tuning on the expected reward rate, it may be that setting a single threshold is effectively closer to "optimal" than setting multiple thresholds. We note that, even though subjects might simply have sped up over the course of early sessions, they appeared to converge on a reward maximizing single threshold on average after six sessions. No further speeding followed this convergence, despite the possibility of doing so, as indicated by faster signal detection response times than decision making response times at asymptote.

Suboptimal performance in economic decision making tasks has previously been explained by the hypothesis that participants favor accuracy above and beyond its contribution to the rewards that can be earned (Maddox & Bohil, 1998). Our results suggest that while participants started the experiment placing a strong emphasis on accuracy, they decreased or eliminated this emphasis with further training to the point that performance was almost exclusively driven by reward rate maximization. Interestingly, this process interacted with the signal quality. Suboptimal performance continued even after extensive training for the difficult and no signal conditions (4 and 0% coherence, respectively), whereas it almost disappeared for the higher signal qualities irrespective of the experienced error proportion. Corroborating this finding, Bogacz et al. (2010) found closer to optimal performance for the easier conditions in another 2AFC task, in which the participants decided if more cells were filled or empty on a 10 × 10 grid. Evidence for optimal performance in a dot-motion task like the one investigated here (Simen et al., 2009) was also based primarily on data that involved accuracies above 85%.

Initially large but diminishing deviations from the optimal speed–accuracy trade-off might be the result of a

long-term strategy that participants adopt when faced with a wide range of signal qualities in a novel stimulus. Specifically, in such instances, participants might initially choose to increase the duration of stimulus viewing on each trial in order to form reliable estimates of crucial task parameters, such as stimulus–response associations, signal-to-noise ratios, distribution of signal-to-noise ratios, reward magnitudes, etc. Reducing uncertainty in these estimations would in the long run benefit participants, allowing them to base their decisions on more accurate and precise task related information. The adoption of this strategy is particularly likely in our study since participants were aware of the multiple-session testing and that the experimental manipulation was the signal-to-noise ratio of a novel stimulus. Increasing drift rate estimates over the course of training support this possible source of suboptimal performance. It should also be expected that the acquisition process would take longer for the low signal-to-noise ratio and no signal conditions (depending on the perceptual ability of the participant). This becomes clearer particularly in the case of the no-signal condition, since it requires the participant to learn that the stimulus contains no information, a process that would be slowed by any prior assumption that all blocks contain motion-direction information that could be detected with more training and vigilance. Informing participants explicitly about the presence of 0% coherence conditions (which we did not do) would address this particular possibility.

Our results suggest a general explanation of the findings of other researchers, in which an increase in speed occurs as a function of practice without any accompanying change in accuracy (e.g., Dye et al., 2009). Perhaps participants in most perceptual decision making tasks are generally converging on the optimum of a reward function as they practice; and perhaps total reward maximization in most tasks favors first tuning up a participant's bottom-up perceptual parameters (by favoring accuracy at the expense of time and short-term reward rate) before fine-tuning such top-down strategic parameters as decision thresholds.

It is important to note that response time distributions were sometimes bimodal, particularly in the difficult task conditions. The shorter mode of these mixture distributions was well-aligned with the mode of the signal detection time distribution. This observation implies that both integrative and non-integrative decision processes were in operation and that participants switched between these two decision strategies (whether or not to accumulate motion direction evidence). Responding at the signal detection latency (i.e. DT = 0) indeed constitutes the optimal decision strategy in the absence of a signal (0%). The interplay between these two decision strategies was further reflected in the decreasing proportion of non-integrative response times with increasing signal quality. In other words, when the stimulus contained more signal, participants were more likely to choose to integrate the evidence and vice versa. Simen et al. (2009) also reported bimodal response time distributions when the probabilities of the leftward and rightward motion directions were unequal. In parallel to our observations, the shorter mode of those distributions was also well-aligned with the signal detection time and its mass diminished as the left/right probabilities approached $p = .5$, corresponding to the reward maximizing strategies under those task conditions. Briefly, these observations suggest that in the 0 and 4% coherence conditions, participants might be switching between optimal and suboptimal strategies, possibly driven by their signal-to-noise ratio uncertainty (particularly given the wide range of experienced signal-to-noise ratios).

Importantly, our results also indicated a reliable relation between temporal uncertainty and deviations from optimal performance. Specifically, participants with higher timing uncertainty about the response-to-stimulus interval were found to deviate more from the optimal performance curve. This suggests another source of suboptimal performance, which was predicted by the asymmetrical relation between reward rates and threshold (decision boundaries) originally reported in the theoretical work of Bogacz et al. (2006) and further analyzed using information gap decision theory in Zacksenhouse et al. (2010). This relation states that if decision makers are to minimize loss due to timing uncertainty in reward rate estimates (uncertainty regarding the denominator of Eq. 2), they should err towards longer than optimal instead of shorter than optimal decision boundaries. Consequently, apparently conservative decision-boundaries might be viewed as an intrinsically adaptive bias in response to timing uncertainty. We note that a similar bias is also predicted by uncertainty in keeping track of average accuracy (the numerator of Eq. 2). Since we have no independent measure of this ER uncertainty, we could not test this hypothesis. There may be additional reasons to favor high decision boundaries as well; namely, to prevent anticipations, which are bound to happen if there is trial-to-trial start point or decision threshold variability.

In line with our findings with respect to optimal performance, recently Starns and Ratcliff (2010) showed that young but not old adults can nearly optimally adapt decision thresholds after extensive training. This age difference might also be partially mediated by differences in the timing abilities of young and old adults; older people show poorer interval timing ability in dual task paradigms (for review, see Balci, Meck, Moore, & Brunner, 2009).

Overall, our findings suggest that two independent factors lead to deviations from optimal performance in

2AFC tasks with response-terminated stimuli. One of these factors appears to be a competition between reward maximization and accuracy that may interact with signal quality (Maddox & Bohil, 1998). This might be an adaptive strategy in the long run, since it could allow participants to form more accurate estimates of the critical task parameters early in training. In particular, it could allow for better estimates of the average signal-to-noise ratio in a task. Since taking extra time to compute better estimates in the service of maximizing long-term rewards would eventually lead to diminishing returns, a gradual shift might be expected from the exploration of task-parameters toward the exploitation of current knowledge for short-term gain. The second process contributing to suboptimal performance appears to be mediated by timing uncertainty, which may be closely linked to a participant's precision in estimating reward rates. Our results show that such a bias might indeed be driven by a participant's level of temporal uncertainty. Finally, it may be that the first process—computing better estimates of signal-to-noise ratios—depends intimately on the second: keeping track of how long decisions are taking. Comparing the accuracy achieved to the time invested in decisions is one way to estimate signal-to-noise ratios, and the current experiment may be taken as evidence in favor such an approach.

Finally, it is also possible that participants start the experiment with a prior that errors result in negative costs, which should disappear when non-negative costs are experienced throughout training. When negative costs are assumed for errors, one would expect the decision times to follow an optimal performance curve that is parameterized by the relative cost assumed for errors (see Fig. 1). As this prior disappears the decision times would converge on the optimal performance curve of the DDM with $q = 0$ approximately following the sequence of curves with decreasing q depicted in Fig. 1. This issue can be empirically addressed by experiments that test both negative and non-negative costs for errors. The relationship between temporal uncertainty and free-response, two-alternative forced-choice performance might be more pronounced in such experiments, where within-participant manipulations include the response-to-stimulus intervals. Response-to-stimulus interval, after all, is the task parameter most directly related to timing capability. Furthermore, the use of timing tasks that are less vulnerable to additional non-timing related sources of variability (e.g., an interval bisection task; Balci & Gallistel, 2006; Wearden, 1991), would add to the generalizability of our findings. We are currently conducting such experiments.

References

Balci, F., Freestone, D., & Gallistel, C. R. (2009). Risk assessment in man and mouse. *Proceedings of the National Academy of Science, 106*, 2459–2463.

Balci, F., Meck, W., Moore, H., & Brunner, D. (2009). Timing deficits in aging and neuropathology. In J. L. Bizon & A. G. Woods (Eds.), *Animal models of human cognitive aging* (pp. 161–201). Totowa: Humana Press.

Balci, F., & Gallistel, C. R. (2006). Cross-domain transfer of quantitative discriminations: Is it all a matter of proportion? *Psychonomic Bulletin & Review, 13*, 636–642.

Barnard, G. A. (1946). Sequential tests in industrial statistics. *Journal of the Royal Statistical Society Supplement, 8*, 1–26.

Bogacz, R., Hu, P. T., Holmes, P., & Cohen, J. D. (2010). Do humans produce the speed-accuracy tradeoff that maximizes reward rate? *The Quarterly Journal of Experimental Psychology, 63*(5), 863–891.

Bogacz, R., Shea-Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *Psychological Review, 113*, 700–765.

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

Dye, M., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science, 18*, 312–326.

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review, 16*, 1026–1036.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84*, 279–325.

Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions and reward. *Neuron, 36*, 299–308.

Holmes, P., Shea-Brown, E., Moehlis, J., Bogacz, R., Gao, J., Aston-Jones, G., et al. (2005). Optimal decisions: From neural spikes, through stochastic differential equations to behavior. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Science, E88A*(10), 2496–2503.

Maddox, W. T., & Bohil, C. J. (1998). Base-rate and payoff effects inmultidimensional perceptual categorization. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 24*, 1459–1482.

Pachella, R. G. (1974). The interpretation of reaction time in informationprocessing research. In B. Kantowitz (Ed.), *Human information processing: Tutorial in performance and recognition* (pp. 41–82). Hillsdale: Erlbaum.

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

Rakitin, B. C., Hinton, S. C., Penney, T. B., Malapani, C., Gibbon, J., & Meck, W. H. (1998). Scalar expectancy theory and peak-interval timing in humans. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 1–19.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 83*, 59–108.

Ratcliff, R. (2001). Diffusion and random walk processes. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 6, pp. 3668–3673). Oxford: Elsevier.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*, 873–922.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*, 347–356.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261–300.

Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., & Cohen, J. D. (2009). Reward-rate optimization in two-alternative decision making: Empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1865–1897.

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging, 25*, 377–390.

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods, 40*, 61–72.

Wald, A. (1947). *Sequential analysis*. New York: Wiley.

Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics, 19*, 326–339.

Wearden, J. H. (1991). Human performance on an analogue of an interval bisection task. *The Quarterly Journal of Experimental Psychology. B: Comparative and Physiological Psychology, 43*, 59–81.

Wickelgren, W. A. (1977). Speed–accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*, 67–85.

Zacksenhouse, M., Holmes, P., & Bogacz, R. (2010). Robust versus optimal strategies for two-alternative forced choice tasks. *Journal of Mathematical Psychology, 54*, 230–245.