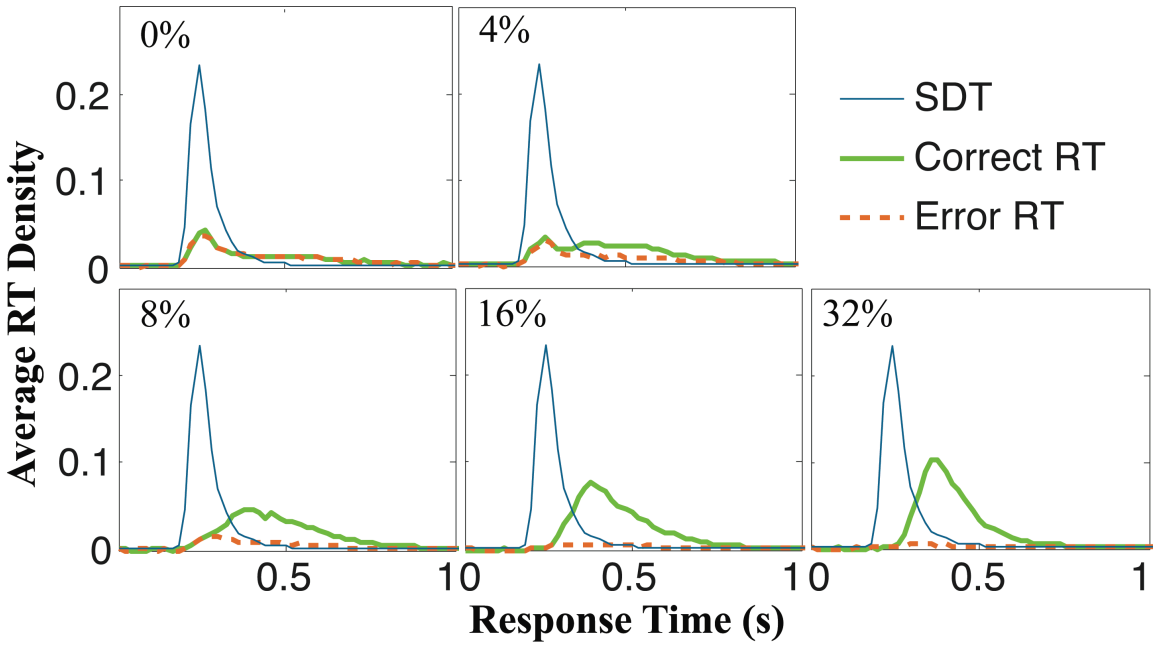
SUPPLEMENTAL MATERIAL

Response time and signal detection time distributions



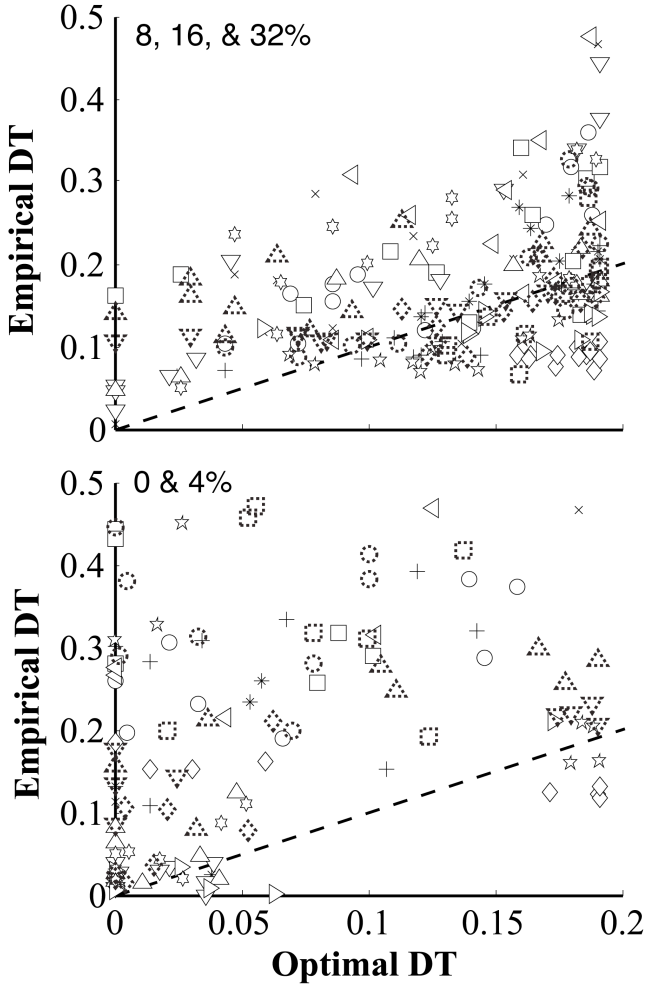
SM Fig. 1. Correct response time (*thick solid green curve*) and error response time densities (*dashed red curve*), averaged across participants, along with average signal detection time density (*thin solid blue curve*). Correct and error response times are normalized by the total number of decisions, and signal detection times are normalized by the total number of signal detection responses. Data were gathered from sessions 10 and above. Note the diminishing mass of error response times with increasing signal quality reflecting improvement in accuracy. Signal detection distributions (SDT) are identical across all five panels

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 (SM Fig. 1). Bimodality was also present in around half of 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. Thus, the optimality analyses for DDM fits to the data collected from 0 and 4% conditions should be interpreted with caution.

Orthogonal regression

In order to assess the degree of match between the empirical and optimal decision-times, we fit an orthogonal regression model (minimizing the perpendicular distances from the data points to the fitted line) in order to account for variability in both the observed and optimal decision time estimates. Slopes of orthogonal regression lines between empirical and optimal decision times for the data collected from sessions 10–13 and from 8, 16, and 32 % coherences were significantly different from 0 [$t(16) = 4.45$, $p < .001$, mean 1.7 ± SE 0.4] but were not significantly different from 1 [$t(16) = 1.89$]. This finding suggests that empirical decision times for the highest coherences tracked the optimal decision times for each error proportion, rather than simply taking a single, fixed value that was close to the optimal performance curve but independent of error proportion (see supplemental material Fig. 2a). In line with our previous analyses, visual inspection of supplemental material Fig. 2b shows that this was not the case for the 0 and 4% coherences.



SM Fig. 2. Empirical decision time as a function of optimal decision time separately for 8–32% and 0–4% coherences (sessions 10–13). Each *symbol* corresponds to data collected from a single participant. Each *data point* corresponds to a block. The identity line is shown *dashed*

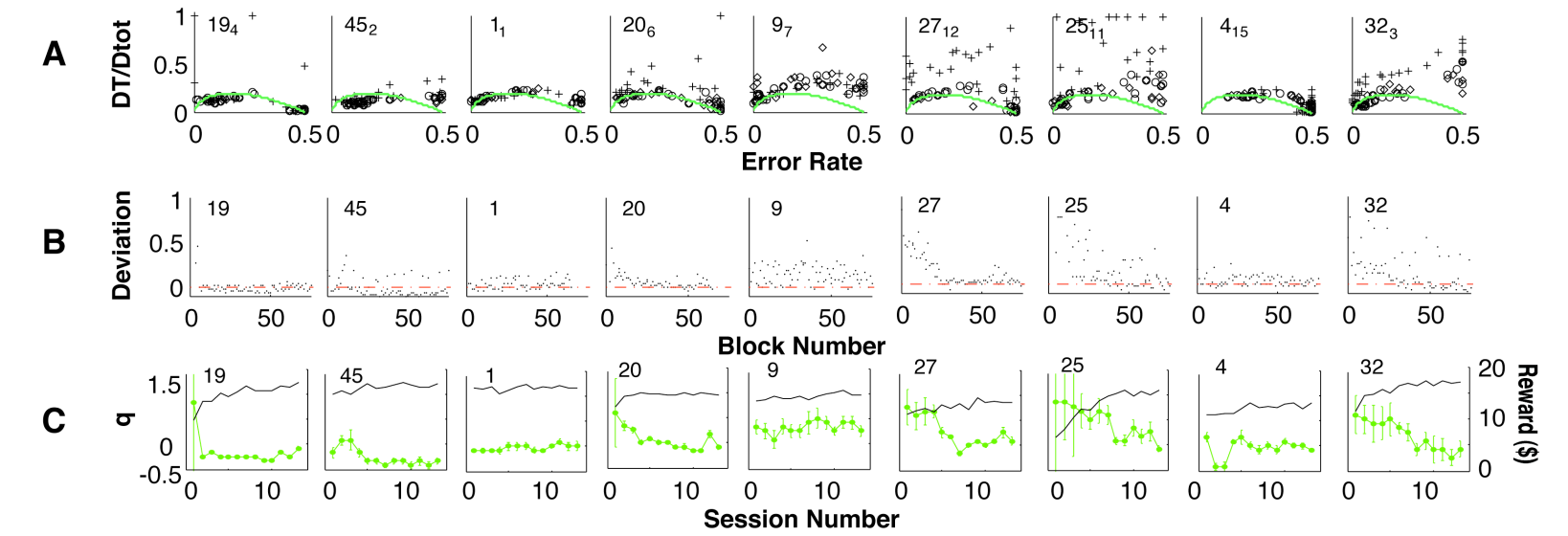
Individual participants' performance

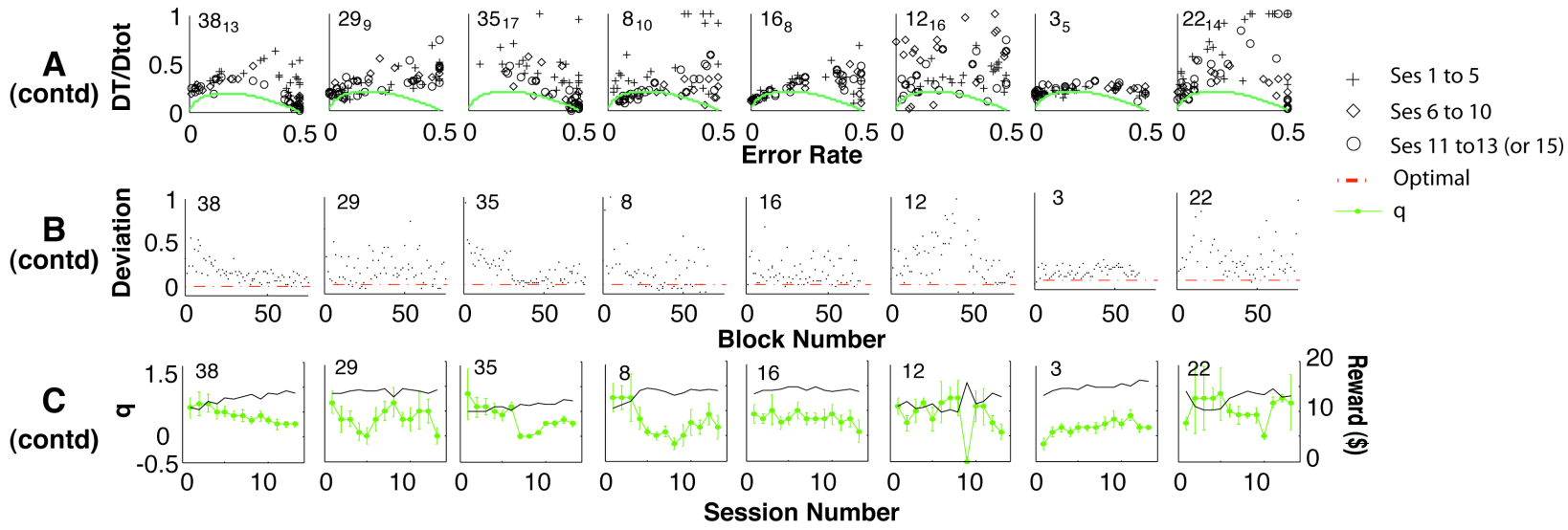
SM Fig. 3a shows normalized decision times as a function of error proportion (ER) across all coherences and for the first 5, second 5, and last 3–5 sessions for individual participants:

$$Normalized\ DT = \frac{RT - T_0}{RSI + T_0} = \frac{DT}{D_{Total}}.$$

The optimal performance curve for the pure drift-diffusion model (DDM) ($q = 0$) is superimposed. Decision time was computed by subtracting an estimate of non-decision time (T_0) from the observed response time (RT). This estimate was obtained from the average response time in the signal detection blocks.

Performance was defined as the mean squared deviation of the fitted decision threshold from the optimal decision threshold, which was calculated from each participant's estimated drift rate (see main text). These values were calculated separately for different coherence levels and participants were sorted in ascending order of their average steady-state performances. Note that this metric is different from the vertical distance between the empirical decision times and the optimal performance curve, information that is implicit in SM Fig. 3. Visual inspection of these plots suggests that some participants were close to optimal from the beginning, some converged on the optimal performance curve over the course of training, and some were asymptotically suboptimal. SM Fig. 3b presents a clearer demonstration of these changes by showing the block by block deviations from the optimal performance curve. Visual inspection of individual plots suggests that more than half of the participants abruptly or gradually converged on the optimal value.

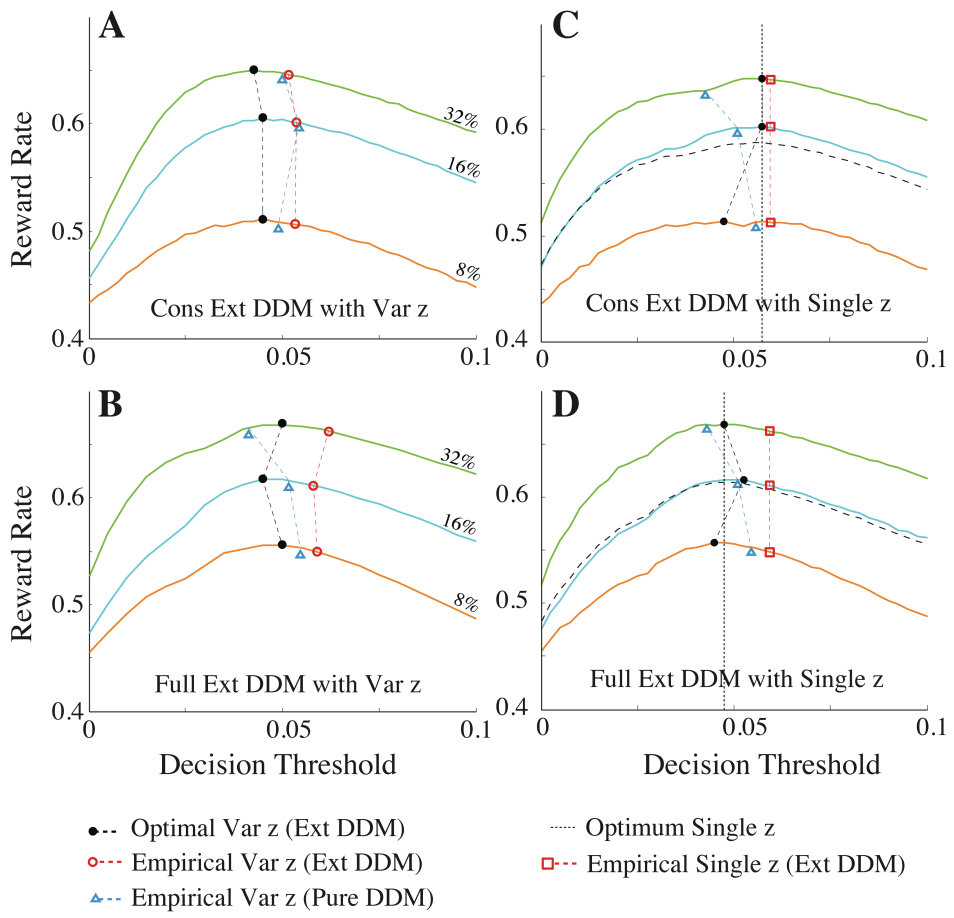

A
DT/Dtot
Error Rate
B
Deviation
Block Number
C
q
Reward ($)
Session Number


A (contd)
DT/Dtot
Error Rate
B (contd)
Deviation
Block Number
C (contd)
q
Reward ($)
Session Number
Ses 1 to 5
Ses 6 to 10
Ses 11 to13 (or 15)
Optimal
q

SM Fig. 3. Each *column* (with three *rows*) corresponds to a different participant. Participant IDs are presented for the first row of each column. The overall/averaged reward rate ranking of the participants is presented as the subscript to the participant ID. From left to right (starting from the *first row of top panel*) participants are ordered based on their average squared deviations from the optimal decision threshold (for sessions 10 and up). **a** Decision time as a function of error proportion presented along with the optimal performance curve of the pure DDM, which is the *skewed inverted-U shaped curve*. Data have been categorized in three groups: first 5 sessions (*crosses*), second 5 sessions (*diamonds*) and final 3–5 sessions (*circles*). **b** Block-by-block deviations (vertical distance) from the optimal performance curve as a function of block number. The *dashed flat line* in this row represents zero deviation from the optimal performance curve of the pure DDM. Decision times and deviations that were higher than 1 have been replaced by 1 for demonstration purposes. **c** Session by session q values for the RR_m fits (*solid green lines with circles*), superimposed on the corresponding monetary rewards (*thin solid black lines*). Participants 1, 3, and 22 had limited previous experience with the dot motion discrimination task

Optimality analysis for the extended DDM

In order to find the optimal threshold for the extended DDM, we simulated the extended DDM with best fitting model parameters (allowing T_0 to vary) and computed the reward rate for different threshold values for the highest three coherences (SM Fig. 4).



SM Fig. 4. Reward rate curves along with empirical and optimal thresholds (*z*) for different extended DDM fits. We fit four different models: **a** constrained and **b** full extended DDM with varied thresholds, and **c** constrained and **d** full extended DDM with a single threshold. Models were fit to the highest three coherences only, with pooled data from multiple participants. On each curve we show the reward maximizing thresholds (*black dashed lines with filled circles*) along with the best fitting threshold values (*red dashed lines with open circles*). On the same curves, we also show the optimal threshold under the pure DDM given the best fitting drift rate (*blue dashed line with open triangles*). Optimal single thresholds are indicated by a *dotted vertical line* in (**c**) and (**d**) whereas the best fitting single thresholds are indicated by the *red dashed line with open squares*. *Dashed curves* in (**c**) and (**d**) are the reward rate curves averaged across the highest three coherences. *Cons* Constrained. *Var* Varied.

SM Fig. 4 shows that for multiple threshold models (a and b), the empirical thresholds track the changes in the optimal thresholds with a fairly constant positive discrepancy. The modulation of the optimal thresholds as a function of coherence in SM Fig. 4a followed a pattern similar to what we observed with pure DDM fits. Note that in

Fig. 4b, in comparison to the pure DDM and constrained extended DDM, the optimal thresholds followed a different pattern as a function of coherence under the fully extended DDM. The best fitting parameter values revealed that allowing a bias in the starting point inflated the starting point variability estimate, which makes the resultant core parameters suspect and not as reliable for optimality analysis (see Simen et al, 2009, for further evidence of parameter inflation in the unconstrained extended DDM). Moreover, there is no task-dependent rationale for assuming a bias in the starting point. Consequently, we take the constrained extended DDM as a more reliable reference for optimality analysis.

SM Fig. 4c, d show that, for the constrained extended DDM, the empirical single threshold was very close to the optimal single threshold; it diverged from the optimal single threshold with the full extended DDM model. When performance was characterized as the proportion of the earned reward rate to the maximum expected reward rate, these proportions were all over 98% for the pure DDM, the extended DDM, and the constrained extended DDM.

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, $(RR_{emp}–RR_{z=0})/(RR_{max}–RR_{z=0})$. When this value was computed in a conservative fashion, the proportion of the maximum expected reward rate ranged between 92 and 99% depending on the fit type. Overall, participants appear to have also performed nearly optimally under the extended DDM irrespective of varied or fixed thresholds.

Issues regarding model fitting

1. We note that the extended diffusion model of Ratcliff & Rouder (1998) includes additional parameters for trial-to-trial variability in starting point and drift rate. Optimal performance in tasks with coherence and mean response-to-stimulus interval held constant within blocks is described by the sequential probability ratio test (SPRT), which in turn requires setting these extra variability parameters to 0. With fewer parameters to fit, this restricted model (which we refer to as the 'pure DDM') leads to higher fit errors and predictions of equal error and correct response time on average. In practice, we find that it also leads to lower fitted values of parameters such as drift and T_0 (cf. Simen et al., 2009).

2. It is important to note that, for the pure DDM, the optimal threshold is 0 in the 0% coherence condition. But without trial-to-trial variability in the non-decision latency, T_0, this implies that response times are totally deterministic, and equal to T_0. This highly implausible assumption must be relaxed to allow for variability in signal detection time—a simpler response process which itself may be modeled as a drift diffusion process on a faster time scale than the typical decision making process. If this assumption is not relaxed, then model fits will be forced to find a way to account for variable response times. This can only be accommodated within the pure DDM by allowing boundaries to be strictly positive. In fact, drift is always 0 in fits of data from the 0% coherence condition, thereby accounting for chance performance. Zero drift implies that the ratio of the fitted threshold to the noise alone determines the spread of the response time distribution. Thus, we should be not be surprised that thresholds, as estimated for the pure DDM, are larger than expected for 0% motion coherence, but the fact that some

participants were on the optimal performance curve (see SM Fig. 3) at 0% motion coherence suggests that their thresholds were in fact optimal for low coherence conditions: when signal-to-noise ratios were particularly low, some participants were able to transition into a form of fast-guess/non-integrative responding (cf. Simen et al., 2009) in which no evidence was accumulated for leftward vs. rightward motion.

Model comparison statistics for single versus multiple thresholds

Table 1 summarizes the model comparison statistics for DDM fits with multiple thresholds vs. a single threshold. Overall, the model with multiple thresholds (two extra free parameters for three coherences) fit the data equally well or better than the model with a fixed threshold.

Table 1. Model Comparison Statistics: Multiple vs. Single Decision Thresholds

	AIC_c (Δ = Multiple *z*–Single *z*)	**BIC (Δ = Multiple *z*–Single *z*)**
Training Period	Median	Median
Session 1	−2.57	5.05
Sessions 2–5	−13.91	−2.64
Sessions 6–9	−12.36	−0.72
Sessions 10–end	−16.62	−4.23

Note. Negative values lend support for the model with multiple thresholds. Models were fit to individual participants' data. *z*: decision threshold

Proportion of maximum possible expected reward rate

Table 2 summarizes the proportion of maximum possible expected reward rate separately for different coherences and fit types. 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 (obtaining 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 these possibilities from skewing our assessment of how closely participants approached optimal performance, we also computed the proportion of maximum earnings in a conservative manner. This latter computation was described above.

Table 2. Proportions of maximum possible expected reward rates during sessions 10 and above

Liberal Definition							
		0%	4%	8%	16%	32%	**Average RR**
0–32% (Multiple *z*)	Mean (SEM)	.78 (.03)	.90 (.02)	.97 (.01)	.98 (.01)	.98 (.01)	***a***
8–32% (Single *z*)	Mean (SEM)	*b*	*b*	.96 (.02)	.99 (.01)	.99 (.00)	**.98 (.01)**
Conservative Definition							
0–32% (Multiple *z*)	Median	*c*	*c*	.97	.98	.99	***a***
8–32% (Single *z*)	Median	*b*	*b*	.96	.99	.99	**.98**
Average		.78	.90	.97	.99	.99	**.98**

Note. Multiple *z*: values computed for independent decision thresholds. Single *z*: values computed for the best fitting single threshold. The measures denoted by *a, b,* and *c* were not computed for the following reasons: (*a)* multiple thresholds were evaluated, (*b)* data from corresponding coherences were not fit with single threshold, and (*c)* the meaning of conservative estimates is unclear when optimal thresholds are very small. Under the same rationale, we used medians rather than means in presenting the output of the conservative analysis. Note that in this particular table the proportions of the maximum expected reward rate when computed conservatively can be greater than those computed under the liberal definition, since medians rather than means were used for the earlier definition. When we used medians for both definitions, as expected, conservative proportions were always smaller than the liberal ones. For the description of different terms refer to main text. RR: Reward Rate

Accounting for deviations from optimality: temporal uncertainty

We also examined the relation between temporal uncertainty and normalized deviations from the single optimum threshold, rather than from the optimal performance curve as

performed in the main text. In the first analysis, we adopted a conservative approach and first entered the *q* estimates from model fits as the primary predictor of the normalized deviations of fitted thresholds from the single optimal threshold (excluding one outlier 2SD above the mean), 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 = .68). ANOVA revealed that the overall model (*q* and CV) was significant, $F(2,13) = 13.59$, $p < .001$. Adding the CV accounted for additional variance, $\Delta R^2 = .13$. When we first entered the CV as the predictor (R^2 = .62, $F(1,14)=22.57$, $p <$ .001), adding *q* estimates improved the fit marginally, $\Delta R^2 = .06$, which was not a statistically significant improvement ($p = .15$).

REFERENCES:

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

Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P. and 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.