

Supplemental Information

The Timescale of Perceptual Evidence

Integration Can Be Adapted

to the Environment

Ori Ossmy, Rani Moran, Thomas Pfeffer, Konstantinos Tsetsos, Marius Usher, and Tobias H. Donner

Supplemental Inventory

The Supplementary Information contains the following three sections:

Supplemental Data

Supplemental Experimental Procedures

Supplemental References.

The section Supplemental Data contains the following items:

- Figure S1 and caption: This figure is associated with main Figure 1.
- Figure S2 and caption: This figure is associated with main Figure 2.
- Table S1: This table is associated with main Figure 2.
- Table S2: This table is associated with main Figure 2.
- Table S3: This table is associated with main Figure 2.
- Figure S3 and caption: This figure is associated with main Figure 3.

1. Supplemental Data

Figure S1, related to Figure 1: Detection performance of leaky integrators with different time constants.

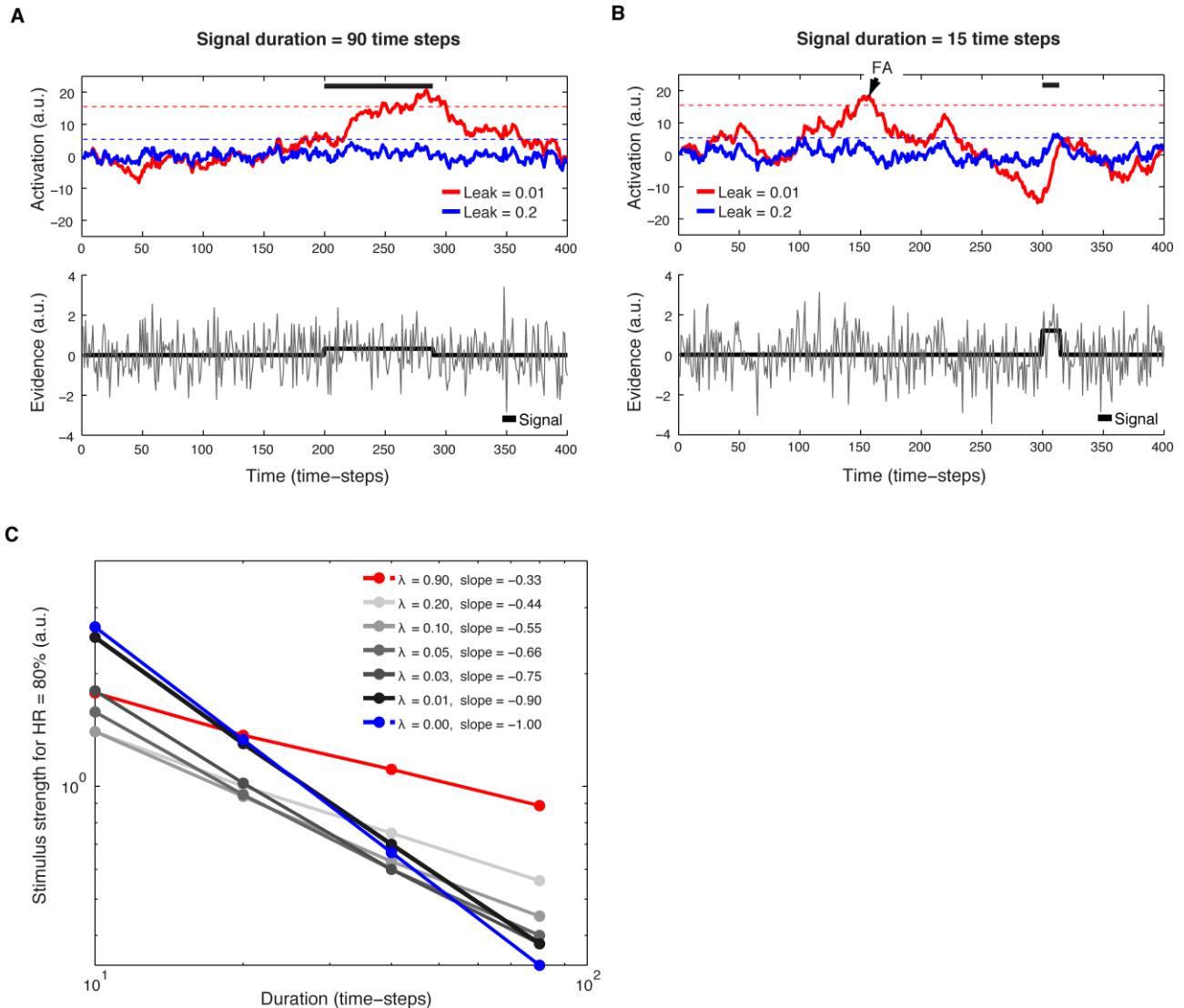


Figure S1 | A. The top panel shows the activation of the integrators with long (red) and short (blue) time constants to the long duration signal shown in the bottom panel. The black line in the bottom panel corresponds to the mean of the signal (boxcar function), to which noise is added to yield the noisy “evidence” represented by the gray trace. The red and blue trajectories in the top panel correspond to the outputs of two leaky integrators in response to the evidence. Here, and in Figure 1, S2 and S3, the model dynamics are as follows: $y(t) = y(t-1) * (1-\text{leak}) + \text{input}(t)$, where the input corresponds to the evidence, and the y-variable corresponds to the output of an integrator with a leak of approximately $1/\tau$ (precisely: $\tau = -10 \text{ ms} / \log(1 - \text{leak})$). A response is elicited when the integrator output crosses a critical “decision bound”, indicated by the dashed lines in the upper panel. In the simulations for Figures 1, and S1-S3, the bound was adjusted separately for the different leak levels, such as to maintain a fixed false alarm rate (20%). Therefore, the red and blue dashed lines (bounds for small and short) are different. The black horizontal bar in the top panel indicates the signal time period. Critically, the long time-scale integrator (red) detects the signal (i.e., crosses the red bound), whereas the short time-scale integrator misses the signal (i.e., does not cross the blue bound). **B.** Same format and model parameters as in Figure S1. Critically, now it is the short time-scale integrator, which detects the signal, whereas the long time-scale integrator

misses it (compare to Figure S1). Instead, the long time-scale integrator generates a false alarm (FA, small black arrow; i.e., bound crossing in the absence of a signal), due to the integration of noise. For low leak values (red line), or in the absence of any leak, integration of the Gaussian noise generates a “random walk”, with a mean of zero, and a variance that increases with time. This may cause crossing of the decision bound due to noise, yielding FAs. The bigger the leak (e.g., blue line), the smaller is the variance of the random walk, and thus the smaller FA. **C.** Detection thresholds of eight integrators with different levels of information leak are plotted as a function of signal duration in log-log coordinates. For each integrator, the response criterion is set so as to maintain a fixed false alarm rate of 20%. The smallest leak ($\lambda = 0$) corresponds to a perfect integrator (blue dashed line), yielding a slope of -1. That is, the detection threshold is inversely proportional to the signal duration: if the signal strength increases by some factor and the duration decreases by the same factor, detection performance stays the same. In other words, the performance of the perfect integrator satisfies Bloch's law, according to which detection is only a function of the 'total signal content'. The opposite extreme occurs for total leak (see red line for a close approximation): the signal information is lost from one time frame to the next. Given a signal of n -frames duration, the observer attempts n independent times (on each frame) to detect the signal and hence improves according to probability summation [1]. Thus, when duration is increased, the detection threshold still decreases, but at a smaller rate than for true signal integration. Intermediate leak values between the previous two extremes produce intermediate slopes (gray lines). In such cases, increasing the duration by a factor, results in a partial compensation for a reduction of the signal by the same factor, due to the fact that some of the signal decays. In sum, the degree of leak determines the slope of the threshold vs. duration function.

Table S1, related to Figure 2: Occurrence of the four error types.

	S-Sessions					L-Sessions				
	Mis-localization	Pre-mature	Too slow	Miss	Total	Mis-localization	Pre-mature	Too slow	Miss	Total
S1	2	32	0	166	200	3	32	3	123	161
S2	1	8	0	180	189	2	15	4	168	189
S3	3	46	1	191	241	2	29	10	160	201
S4	1	63	0	158	222	4	27	0	113	144
S5	0	18	0	210	228	2	10	2	163	177
S6	3	14	0	308	325	1	15	0	231	247
S7	3	39	0	183	225	2	23	1	174	200
S8	2	45	0	304	351	1	32	0	250	283
S9	2	7	9	347	365	1	6	4	345	356
S10	0	64	1	174	239	0	16	1	142	159
S11	0	34	0	252	286	0	41	0	217	258
S12	0	73	0	109	182	1	28	1	171	201
Total	17	443	11	2582	3053	19	274	26	2257	2576

Table S1 | Numbers of errors per error category, observer, and session type (S for short dominance and L for long). “Mis-localization”, response on signal+noise trials within response window but with incorrect button; “pre-mature”, response on signal+noise trials before response window; “slow”, response on signal+noise trials after the response window; “miss”, no response during signal+noise trials.

Figure S2, related to Figure 2: Computational model fits support adaptive temporal integration account.

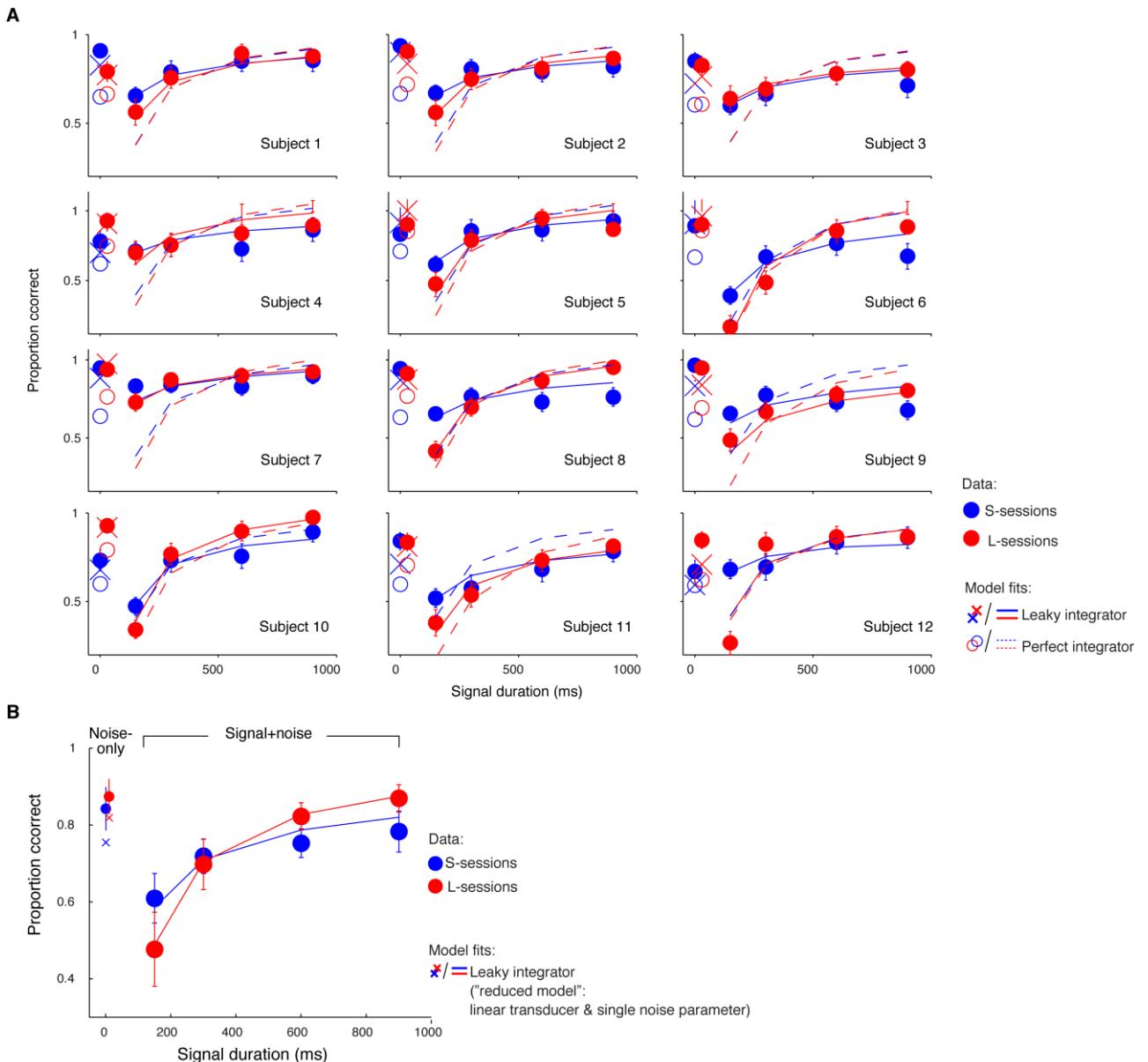


Figure S2 | A. Individual predictions of best-fitting leaky integrator model (solid lines) with separate leak parameters for L (blue)- and S (red)-sessions. See *Supplemental Experimental Procedures* for details. Best fitting diffusion model (dashed lines) are plotted together with the individual proportion correct (correct rejection rate and hit rate) data for all nine observers. Error bars, 95% confidence intervals. Note that, in three observers (3, 6, and 9) hit rates did not increase from 600 to 900 ms signal in the S-sessions (compare two rightmost blue data points in these three observers, exhibiting a statistically non-significant trend towards a drop). Even probability summation of signal detection events over time (i.e., no explicit mechanistic temporal integration) predicts an increase in hit rates with signal duration [1]. It is possible that this idiosyncratic feature reflects the use of a different decision strategy in the S-sessions: These observers might base their decision, in addition to the integrated signal, on neural transients evoked by signal on- and offsets in the S-sessions. This may explain why hit rates do not improve with signal duration. **B.** The group average ($N = 12$) predictions of best-fitting leaky integrator model with simplified parameters (“reduced model”): As in Figures 3 and S5, except a linear transformation of signal intensity into input to integrator (“linear transducer”, no exponent for input non-linearity) and a single noise parameter for both L- and S-sessions. Error bars, 95% confidence intervals. See *Experimental Procedures* for details. The

results show that the simplified model variant accounts for the effect of session-type on the psychometric function, similarly to the full model presented in Figure 3: Performance for short and long signal durations is better in the S and L sessions respectively, a result of integrating with a larger time constant in the L session. Qualitatively identical results were obtained with two additional model variants (data not shown): (i) A linear transducer variant in which integration noise was free to vary across session type; (ii) A variant with a single noise parameter across session types and non-linear transduction. In sum, computational modeling of the decision process supports the adaptive leaky-integration account we propose here. This conclusion does not depend on our assumptions about variability of integration noise across session time and non-linear transduction.

Perfect integration provides the optimal algorithm for signal integration of signals of constant intensity, when there is no temporal uncertainty. However, under conditions of temporal uncertainty this integration strategy also perfectly integrates the noisy samples prior to the signal onset. Thus, (i) more premature responses (prior to signal onset) ensue; and (ii) when the signal finally onsets the integrator is already in a very noisy state, which interferes with its ability to detect the signal (Figure S1B, red line, showing integration with long time constant). Unlike perfect integration, when integration is leaky, the impact of noisy samples from the past is attenuated, and thus pre-signal noise affects the integration to a lesser extent. The cost is that signal samples are also subject to 'forgetting' and thus the observer fails to extract the maximal possible SNR from the signal samples. Thus, leak implements a tradeoff between extracting the maximal SNR (during the signal) and reducing pre-signal noise. By integrating evidence with leak under temporal uncertainty the observer can harness, rather than suffer, leakage to her advantage.

Table S2, related to Figure 2: Best-fitting parameters for leaky integrator model.

Observer	B	σ_{in}^L	σ_{in}^S	τ^L	τ^S	θ^L	θ^S
1	0.90	0.16	0.10	28.07	8.58	2.51	1.35
2	0.70	0.14	0.04	17.68	4.75	1.99	0.94
3	0.80	0.10	0.10	4.75	3.83	1.03	0.94
4	0.90	0.14	0.02	16.16	3.48	1.95	0.76
5	0.50	0.12	0.08	32.83	7.18	2.31	1.04
6	0.30	0.10	0.10	99.50	5.74	2.66	0.89
7	1.00	0.06	0.00	6.63	4.48	1.18	0.90
8	1.00	0.16	0.10	66.17	3.04	3.61	0.93
9	0.70	0.22	0.08	6.63	2.92	1.73	0.82
10	0.40	0.00	0.10	99.50	32.83	2.38	1.74
11	0.40	0.18	0.10	13.78	5.37	1.87	0.90
12	0.80	0.08	0.00	199.50	11.26	3.52	1.11
Mean	0.70	0.12	0.07	49.27	7.79	2.23	1.03

Table S2 | See *Experimental Procedures* for explanation of the model parameters. The superscripts S and L correspond to the short and long session types respectively; β , exponent characterizing the individual brightness perception; σ_{in} , integration noise of the leaky integrator; τ , time constant of leaky integrator (number of 10 ms monitor refreshes); θ , decision bound.

Table S3, related to Figure 2: Best-fitting parameters for the perfect integrator (diffusion) model.

Observer	β	σ_{in}^L	σ_{in}^S	θ^L	θ^S
1	1.00	0.00	0.02	4.44	4.37
2	1.00	0.00	0.00	4.72	4.39
3	1.00	0.04	0.04	4.27	4.22
4	1.00	0.02	0.00	4.59	4.09
5	1.00	0.02	0.00	5.10	4.39
6	0.90	0.12	0.12	6.40	5.35
7	1.00	0.00	0.00	4.85	4.19
8	1.00	0.02	0.00	4.83	4.19
9	1.00	0.14	0.00	5.98	4.14
10	1.00	0.00	0.00	5.20	4.09
11	1.00	0.18	0.00	7.02	4.09
12	1.00	0.02	0.00	4.22	4.09
Mean	0.99	0.05	0.02	5.14	4.30

Table S3 | See *Experimental Procedures* for explanation of the model parameters. The superscripts S and L correspond to the short and long session types respectively; β , exponent characterizing the individual brightness perception; σ_{in} , noise of the integrator; θ , decision bound,

Figure S3, related to Figure 3: Threshold vs. duration functions support adaptive temporal integration account.

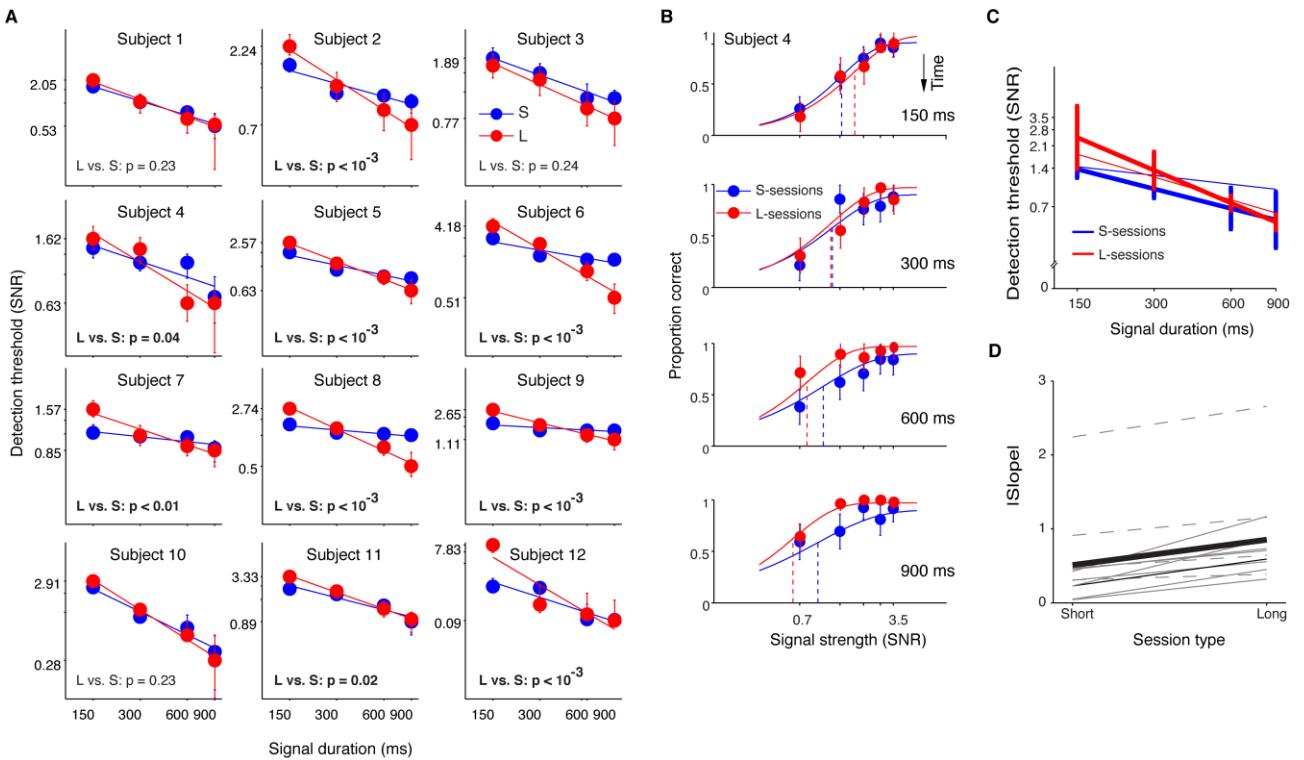


Figure S3 | A. Detection thresholds, plotted as a function of signal duration in log-log coordinates for each individual observer, separately for L- and S-sessions. Conventions as in Figure 4B. Error bars, 95 % (bootstrap) confidence intervals. P-values correspond to a one-sided permutation test of the difference between slopes in L- and S-sessions. Statistically significant p-values are printed in bold. Nine out of twelve subjects show a significant difference in the slopes. **B-D** Threshold vs. duration functions, based on threshold extraction procedure with $\log(\text{threshold})$ vs. $\log(\text{duration})$ linearity constraint. **B.** Psychometric functions of one example observer, shown separately for the four signal durations and two sessions types. Psychophysical detection thresholds (expressed in units of signal-to-noise ratio, SNR) decreased with duration, but more strongly in L- than in S-sessions. The threshold parameter (vertical dashed lines) corresponded to the signal strength eliciting a fixed hit rate within each observer, which ranged from 57% to 63% across observers, depending on the (individually estimated) chance level. The logarithm of the detection threshold was constrained to be linear in the logarithm of signal detection. Thus, for both S- and L-sessions, the threshold decreases equally from signal duration 150 ms to 300 ms, from 300 ms to 600 ms, etc (the logarithmic spread is identical for these durations) **C.** Group average thresholds, plotted on log-log scales as a function of signal duration. Thick lines, group average. Thin lines, observer from panel A. Error bars, 95% confidence limits. **D.** Individual (thin lines) and group average (thick line) best-fitting $\log(\text{threshold})$ - $\log(\text{duration})$ slopes for S- vs. L-sessions. The absolute value of the slopes is shown here (fitted slopes are negative; see panel C). Thin black thin line, observer from panel A. See *Experimental Procedures* for details. At the group level, these slopes changed from 0.51 to 0.85, in S- and L-sessions, respectively. Slopes were significantly larger (i.e., more negative) in the L than S-sessions ($p < 0.05$; permutation test) in eight of the twelve observers (solid gray lines). The difference in slopes was also highly significant when tested for the entire group (Wilcoxon signed rank test; $p < 10^{-3}$). The results confirm the model prediction of higher threshold-duration slopes, in the L- compared with S-session (**Figure 1**).

2. Supplemental Experimental Procedures

Rationale for choice of distributions of signal durations

Multiple different distributions of signal durations would, in principle, allow for manipulating the typical signal duration across sessions. Here, we opted for keeping the numbers of trials constant for all durations, except for the typical one (the shortest or the longest, respectively). This choice was based on the following two considerations. First, our goal was to estimate thresholds as a function of signal duration, while encouraging observers to employ maximally different integration time scales in both types of sessions, provided that such time scale adaptation was possible at all. Obtaining reliable threshold estimates required a minimum number of trials per duration. Obtaining different integration time scales requires introducing a large difference between the longest and the shortest duration. The chosen procedure was well suited to satisfy both requirements. Second, our procedure made it easy to explicitly instruct naive observers about the main strategic aspect of the task, by pointing them to a single predominant duration (as opposed to describing a whole distribution of signal durations).

Data analysis: general

We began all of the procedures described below with data filtration: For each observer, we removed the first block (100 trials) of each session-type (i.e., the first block of the first S and the 1st block of the 1st L-Sessions). We reasoned that, during this block, participants adapted to the distribution of signal durations for that session. Data was not removed for later sessions of each type. We assumed that in these later sessions, participants utilized their previously developed strategy. Our analysis is based on a total of 900 trials per session type for nine observers and on 1400 trials per session type for three observers, respectively.

Fitting and comparing models of the decision process

Model description: leaky integrator with variable time constant

The model dynamics were specified in time-steps of 10 ms (the monitor refresh-time). The model contained two “psychophysical channels”, which correspond to populations of sensory neurons transforming the external stimuli into the brightness perceived by the observer, and, likewise, the input to the decision stage (integrator, see below). The output of each channel is modeled at time frame t , as

$$\text{Channel}_i(t) = (b + \sigma_{\text{stim}} * \dot{\eta}_i(t) + \text{input}_i(t))^\beta \quad \text{Eq. 1}$$

where $b = 0.4$ is the baseline of each channel, $\sigma_{\text{stim}} = 0.11$ is the standard deviation of the Gaussian noise imposed on each channel, $\dot{\eta}_i(t)$ is a random standard (mean= 0, std=1) Gaussian sample (independent across frames and across channels), β is a “psychophysical brightness exponent” (a free parameter of the model explained below) and $\text{input}_i(t)$ is the (luminance) input signal to the channel at time frame. For ‘purely noise’ frames $\text{input}_i(t) = 0$, whereas for signal frames (see below), $\text{input}_i(t)$ is equal to the signal strength i.e. one of the values 0.08, 0.16, 0.24, 0.32, 0.40. For simplification, for signal+noise trials, we modeled the signal onset after 2 seconds from the beginning of the trial, which corresponds to the average signal onset time in the experiment. (We also examined another variant of the model, in which the onset time was uniformly distributed between 0.6 and 3.5 s as in the experiment, with qualitatively identical results.) Therefore the signal channel received non-zero $\text{input}_i(t)$ input samples for t after 2 s and for the signal duration (150, 300, 600, or 900 ms).

Next, we assumed that on each frame the activation difference between the perceptual

channels, defined by $\Delta Ch(t) = Channel_1(t) - Channel_2(t)$, is fed into a leaky integrator $y(t)$. The dynamics of the leaky integrator are governed by the equations:

$$y(t) = y(t-1) * (1 - \text{leak}) + \Delta Ch(t) + \sigma_{in} * \zeta(t) \quad \text{Eq. 2}$$

$$y(0) = 0 \quad \text{Eq. 3}$$

where leak is the integrator's leak parameter, σ_{in} is the standard deviation of an 'internal' noise in the integration process and $\zeta(t)$ is a random standard Gaussian sample independent across t (and independent of $y(t)$). Both leak and σ_{in} are free parameters of the model. Eq. 3 states that on each trial the leaky integrator initiates at 0, with no response bias for either of the two alternatives. The variable y can be either positive or negative, indicating a signal in the right or left location, respectively. Thus a zero-starting point corresponds to a lack of decision bias. This formulation is equivalent with having two integrators, one which accumulates $Channel_1(t) - Channel_2(t)$, and the other $Channel_2(t) - Channel_1(t)$, towards a common response criterion, θ , and whose starting points are equal.

A response is elicited when the absolute activation of the integrator $y(t)$ crosses one of two symmetrical decision bounds, corresponding to each channel, and denoted by θ and $-\theta$ respectively. θ is a free parameter of the model and by convention the positive bound corresponds to the detection of a left-signal, and the negative bound to the detection of a right-signal. We did not model reaction times (RTs). Observers had to respond within 600 ms following signal offset. Therefore, we assumed for simplicity that the motor delay in the model was fixed at 600 ms. Thus, for noise trials a false alarm ensues if and only if the integrator elicits a response by $t = 4.4$ s, which is obtained by subtracting the 600 ms motor response time from the 5 s trial duration. For signal+noise trials a Hit ensues if and only if at the first time the integrator crosses a bound two conditions hold: 1) It crosses the bound that corresponds to the signal's direction (i.e. θ is the signal is presented on the left side and $-\theta$ if the signal is presented on the left side); and 2) the crossing time satisfies $1.4 \leq t \leq 2 + \text{signal duration}$ (in the model the signal onsets at $t = 2$).

The model had the following four free parameters: leak of the integrator, internal Gaussian noise (quantified as the standard deviation σ_{in}), decision bound θ , and the psychophysical brightness exponent β . In the model, we allowed the leak and the decision bound to vary across session types, assuming that these are both strategic parameters which observers may adjust to improve their performance. The psychophysical brightness exponent, on the other hand, characterizes the sensory signal transduction, that is the transformation of luminance input into a perceived brightness value. This exponent was assumed to be a constant psychophysical property of each observer, and it was, therefore, maintained at a fixed level across session types. The integration noise parameter was also free to vary across session types (reflecting random fluctuations from session to session), but we also tried a more constrained version in which it was constrained to be identical across session types (see below). Thus, in total, the leaky integrator model consists of 7 free parameters. In addition we examined more parsimonious sub-models obtained by constraining the brightness exponent to be 1 and/or allowing only a single integration noise parameter; see below).

Model description: Perfect integrator (drift diffusion model)

The only difference between the leaky integration and the drift diffusion model is that in the latter, the integrator isn't subject to leakage. Therefore, the description of the diffusion model is identical to the description provided above with a single change: In Eq. 3, we constrained $\text{leak} = 0$. This reduces the number of free parameters by 2, (the leak for each session type) yielding a model with a total of 5 free parameters.

Model fitting procedures

We fitted both leaky integrator and perfect integrator (drift diffusion) models to each observer's data separately. Our fitting method was based on maximum likelihood (ML) estimation. The model was

fitted to 42 free data points given by the accuracy rates for noise-only trials (1) and for each signal duration (4) X strength (5) combination (for a total of 21 conditions per session type) for each of two session types. Thus, both the fits of the leaky integrator model ($42 - 7 = 35$ degrees of freedom) and of the diffusion model (37) are highly constrained.

First, we defined a search grid for the brightness exponent β , which varied across the values 0.1, 0.2, ..., 1. Next we repeated the following procedure for each β on the grid search (so from now on β is maintained at a fixed level): We constructed a 3 dimensional search grid with the following dimensions: (i) The integration noise σ_{in} which varied across 13 values (0, 0.02, 0.04, ..., 0.24); (ii) the integration leak, $leak$, which varied across 40 values between 0 and 0.3 (from 0 to 0.1 in steps of 0.005 and then from 0.11 to 0.3 in steps of 0.01); (iii) the decision bound θ , for which search values were determined in the following manner.

For each pair of noise-leak parameters we simulated 100,000 noise only trials according to Eq. (3,4) for the whole duration of the trial (save the motor time) i.e. 4.4sec. For each trial we calculated the maximal absolute value the integrator obtains during the trial: $\max_{0 \leq t \leq 4.4} |y(t)|$. We thus obtained the distribution for the maximal absolute integrator value, whose 99,98,..., 60 percentiles give the 40 bound values on our grid search, $\theta_1(leak, \sigma_{in}, \beta)$, $\theta_2(leak, \sigma_{in}, \beta), \dots, \theta_{40}(leak, \sigma_{in}, \beta)$, which yield false alarm rates of 1,2,...,40% respectively. We thus obtained (for a given β) a three-dimensional (3D) search grid with $13 * 40 * 40 = 20800$ parameter combinations.

Next, for each $(leak, \sigma_{in}, \theta_j)$ parameter triplet and session type we calculated the likelihood term. The model predictions for false alarm rate are readily obtained by $\hat{fa}=j$ (the subscript of θ_j). We obtained the model predictions for hit rates by simulating 10,000 signal+noise trials for each signal strength x duration combination and we calculated the hit rates $\hat{h}_{d,s}$, where the subscripts d and s denote the duration and the strength of the signal correspondingly. The likelihood of the parameter triplet is given by:

$$L_{st}(leak, \sigma_{in}, \beta, \theta_j(leak, \sigma_{in}, \beta)) = \\ \binom{N_{noise}}{N_{fa}} \hat{fa}^{N_{fa}} (1 - \hat{fa})^{(N_{noise} - N_{fa})} \times \\ \prod_{d=1}^5 \prod_{s=1}^4 \binom{N_{signal,d,s}}{N_{hit,d,s}} \hat{h}_{d,s}^{N_{hit,d,s}} (1 - \hat{h}_{d,s})^{(N_{signal,d,s} - N_{hit,d,s})} \quad \text{Eq. 2}$$

Where the subscript st denotes the session type (S-'Short' or L-'Long') and N_{noise} , N_{fa} , $N_{signal,d,s}$, $N_{hit,d,s}$ are the numbers of noise, false alarm, signal (of duration s and strength s) and hit (for duration d and strength s) trials respectively (for session type st). Since both session types are independent, the Likelihood of a parameter sextet i.e. a leak-noise-bound per session type, is given by:

$$L^S \left(leak^S, \sigma_{in}^S, \beta, \theta^S \left(leak^S, \sigma_{in}^S, \beta \right) \right) * L^L \left(leak^L, \sigma_{in}^L, \beta, \theta^L \left(leak^L, \sigma_{in}^L, \beta \right) \right) \quad \text{Eq. 3}$$

where the superscripts correspond to the session type. We maximized the term in Eq. 6. by maximizing each term L^S , L^L separately, which was achieved by means of searching on the 3D grid (the only common parameter for these terms is β which is maintained at a fixed level for the moment) and obtained a 'best fitting parameter sextet'. By repeating the above for each value of β ,

we found the maximal likelihood values, conditional on each β . We completed our parameter search by maximizing the likelihood term over β . We adjoined this best β value with the best fitting sextet for β to generate the maximal likelihood septuplet.

Finally, In order to refine the parameters and to extrapolate to potential parameter combinations out of our search grid, we conducted the iterative Nelder-Mead SIMPLEX procedure [2], to minimize the inverse of the logarithm of the likelihood term in Eq. 6 (equivalent to maximizing the likelihood). Our best search-grid parameter septuplet served as the starting point for the simplex search. This follow-up simplex procedure yielded only minor changes in both the calculated (inverse log) likelihood and in the estimated parameters. We, therefore, report the grid search parameters in this paper.

To fit the drift diffusion model we used the same likelihood calculations, but in searching for the ML parameters we considered only grid points that satisfy $\text{leak}^S = \text{leak}^L = 0$, reducing in effect the dimensionality of the search. Thus, for each brightness exponent β we constructed a 2D, rather than a 3D search grid (for noise and bound with no leak). Except for that change we identified the ML parameter quintet $\sigma_{in}^S, \sigma_{in}^L, \theta^S, \theta^L, \beta$ in the exact same manner.

Model comparison

Model comparisons were performed based on the Bayesian information criterion (BIC) [3]. This criterion implements a tradeoff between ‘goodness of fit’, gauged by the inverse of twice the log likelihood value, and model parsimony, measured by the number of free parameters in the model. The criterion thus penalizes models for their complexity so as to test whether the improvement in fits provided by the more complex models justifies their reduced parsimony. When selecting one out of several alternative models, the model with the minimal criterion value is preferred. BIC penalized by $\ln(N)$ per each of the model parameters, where N is the total number of observations per observer:

$$\text{BIC} = -2\ln L + k\ln(N) \quad \text{Eq. 7}$$

We calculated the BIC values for each observer for both the leaky integrator and the diffusion model. Note that $k=7$ or $k=5$ for the leaky integrator and for the drift diffusion sub-model respectively, and $N=1800$ or 2800 for observers 1-6,10-12 and 7-9 respectively (Table 1).

Alternative variants of leaky integrator model

We also tested a more parsimonious variant of the full leaky integrator model reported in the main paper, to ensure that our conclusions were robust with respect to the assumptions that (i) the integration noise may vary across session type, and (ii) that there is a non-linear transformation from stimulus to integrator input (psychophysical brightness exponent). This variant was identical to the full model with the exception that σ_{in} was constrained to be identical across session types, $\sigma_{in}^S = \sigma_{in}^L = \sigma_{in}$, and the brightness exponent was constrained to 1: $\beta=1$, reducing the number of free parameters from seven to five. The fitting procedure was similar, with the following change: For each σ_{in} we constructed a 2D search grid for leak and bound as before, and found the maximal likelihood leak-bound pair for each of the 2 sessions (conditional on σ_{in}). We then maximized these conditional maximal likelihoods over σ_{in} to obtain the best noise parameter. This parameter was adjoined by the best noise-bound pair for each session (conditional on that noise parameter), to yield the ML parameter quintet. This model variant provided highly consistent results supporting our conclusions (compare Figures 2 and S2). We also fitted by similar means, two ‘intermediate’, six-parameter, variants obtained from the full leaky integrator model by relaxing each of assumptions (i) and (ii) separately. These variants too provided highly consistent results, in support of our conclusions and are hence not reported.

Linear fits of threshold vs. duration functions

Fitting cumulative Weibull functions to the psychometric functions

The noise-only trials were only used to estimate each observer's false alarm rate (FA, see below), but were not otherwise used in the analysis. For signal+noise trials, we computed the hit rate as a function of signal strength (S) and fitted a family of cumulative Weibull functions to the fit data from each session type (Quick, 1974):

$$\Psi(S, t) = \delta(t) + (1 - 2\delta(t) - \lambda) \left(1 - \exp \left[-\left(\frac{S}{\alpha(t)} \right)^{\beta(t)} \right] \right) \quad \text{Eq. 8}$$

where S denotes signal strength, t denotes signal duration, the free parameters α , β , and δ denote the threshold, slope, and lower asymptote (i.e., chance level) parameters of the fitted Weibull function, for each signal duration t . The additional parameter δ compensated for premature responses occurring prior to signal presentation and was, therefore, independent of signal duration.

δ was determined as the probability of pressing a button before signal onset, given each observer's false alarm rate (FA). FA served as a proxy of the individual's tendency to randomly press a button in the absence of a signal. The average signal onset time was 2 s, that is 0.4 times the total trial duration of 5 s. Therefore, the probability of a premature response was estimated as $= 0.4 \times \text{FA}$. The chance level (δ in Eq. 8) on signal+noise trials also depended on the observer's tendency to randomly press a button, as estimated by FA. Specifically, δ depended on $0.5 \times \text{FA}$, taking into account a random selection of one of two response buttons (left vs. right). Thus, δ was estimated as $\delta = 0.5 \times \text{FA} \times P_{\text{random}}$, where P_{random} was the probability of hitting the time of a signal, when randomly pressing a response button. P_{random} was computed as the duration of the response window (from signal onset to 600 ms after signal offset), divided by total duration of the trial (5 s).

The threshold parameter α corresponded to the signal strength, which elicited a certain hit rate on signal+noise trials. Since chance level varied slightly across observers (see above), the threshold-level hit rate also varied across observers, ranging from 55% to 62%.

For each session type and all four signal durations, a set of parameters α and β was estimated using a maximum likelihood procedure [4]. Assuming that the data had been generated by a Bernoulli process and that the probability observing a certain proportion of correct responses follows a binomial distribution, we computed the likelihood for a given set of parameter values for each signal duration t , separately for both session types:

$$L(t) = \prod_i \left[\binom{n_i}{y_i n_i} \Psi(S_i, t | \theta)^{y_i n_i} [1 - \Psi(S_i, t | \theta)]^{n_i - y_i n_i} \right] \quad \text{Eq. 9}$$

where n_i denotes the number of trials with a signal strength S_i and y_i is the corresponding proportion of correct responses. Ψ is the hit rate predicted by the model, given a set of parameter values θ . We obtained the estimates of thresholds and slopes for all durations simultaneously by iteratively adjusting the set of 8 free parameters (4 threshold and 4 slope parameters) by means of the Nelder-Mead SIMPLEX procedure [2].

We regularized the model to minimize over-fitting. Both slope and threshold parameters were constrained to decrease monotonically as a function of signal duration. The slope parameter was further constrained to change by maximally a factor of 2. Both constraints were derived from fitting simulated data produced by the leaky integrator model. For one observer (subject 10), performance was at ceiling for the 900 ms duration (L-sessions), yielding a bad fit for this condition. In this case, we allowed the slopes to decrease by a larger factor, because this improved the goodness of fit in this observer without changing the qualitative pattern of results.

Confidence intervals for each data point were obtained from a parametric bootstrap ($N=1000$) with binomial variability [5]. Subsequently, we regressed (separately for each observer and session type) the logarithm of the detection threshold alpha against the logarithm of the signal duration, yielding a regression slope per observer and session type. These slope values were then compared within and across observers (see below).

We also performed an alternative approach for the Weibull fits, which was even more constrained. This approach focused on the hypothesis that the linear slope of the logarithm of the

detection threshold vs. the logarithm of signal duration differs between S vs. L sessions (Figure 1). Thus, we constrained $\log(\alpha)$ to be a linear function of $\log(\text{duration})$. Instead of four free α parameters (one per signal duration), the fit had only two free parameters for each session type: The intercept and slope of this linear relation. This approach also eliminated the necessity for the subsequent linear regression step (see above). We also assumed that $\log(\beta)$ is linear with $\log(\text{duration})$, further reducing the number of free parameters by 2. We allowed β to be free, allowing for attention lapses. β was maintained at a fixed level across durations as was limited to a maximal value of 10%. In total, we had five free parameters per session type: intercept and slope for each of the linear $\log(\alpha)$ and $\log(\beta)$ relations with $\log(\text{duration})$ and β . This fit was also based on maximal likelihood estimation and we used the same SIMPLEX procedure. The results from this procedure were qualitatively identical (see Figure S3, B-D), thus also supporting the conclusions of this paper.

Testing for differences between slopes of threshold vs. duration functions

For both of the alternative psychometric fitting procedures described in the previous section, we used a one-sided permutation test [6] to compare the slopes of the fitted threshold vs. duration functions between session types (S/L), within individual observers (Figures S3A). For each signal strength and duration on signal+noise trials, as well as the noise-only trials, the session labels were randomly rearranged (i.e., the labels were resampled without replacement). Detection thresholds were estimated for the permuted data sets as described above, and threshold differences between the two session types were computed. Repeating this procedure N-times (with N equal to or larger than 3000) yielded a distribution of difference values under the null hypothesis of no difference in slope, against which we compared observed differences. The p-value was obtained by dividing the number of cases, for which the permuted difference value was equal or greater than the actual difference by the number of permutations. For both procedures, we used a Wilcoxon signed rank test to compare the slopes (Figures 3 and S3, B-D) between session types (S/L) across the group.

3. Supplemental References

1. Watson, A.B. (1979). Probability summation over time. *Vision Res* 19, 515-522.
2. Nelder, J.A., and Mead, R. (1965). A simplex method for function minimization. *Comp J* 7, 308-313.
3. Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6, 461-464.
4. Wichmann, F.A., and Hill, N.J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & psychophysics* 63, 1293-1313.
5. Wichmann, F.A., and Hill, N.J. (2001). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & psychophysics* 63, 1314-1329.
6. Efron, B., and Tibshirani, R. (1998). An introduction to the bootstrap, 1st CRC Press reprint. Edition, (Boca Raton, Fla.: Chapman & Hall/CRC).