Supplemental Materials

S1. Adjusting evidence strength criteria to obtain mirror effect in equal variance SDT

In this section, we illustrate how to adjust decision criteria on the strength axis in order to obtain the mirror effect in equal variance signal detection theory (SDT) model. As discussed in the main text, the mirror effect holds when the intercept is higher than 0 for the zHR$_{\text{easy}}$/zHR$_{\text{difficult}}$ plot, and lower than 0 for the zFAR$_{\text{easy}}$/zFAR$_{\text{difficult}}$ plot. In equal variance SDT, zHR and zFAR can be calculated as (see also the main text):

$$zHR = \mu_S - C$$

$$zFAR = \mu_N - C$$

in which μ_S and μ_N are the mean of the signal and noise distribution, respectively, and *C* is the decision criterion on the strength axis. Suppose that the evidence strength criteria in easy condition (C_{easy}) is a function of those in difficult condition ($C_{difficult}$):

$$C_{easy} = f\left(C_{difficult}\right)$$

Then the relationship between zHR$_{\text{easy}}$ and zHR$_{\text{difficult}}$, and between zFAR$_{\text{easy}}$ and zFAR$_{\text{difficult}}$, can be written as:

$$\begin{aligned} zHR_{easy} &= \mu_{S(easy)} - C_{easy} = \mu_{S(easy)} - f\left(C_{difficult}\right) \\ &= \mu_{S(easy)} - f\left(\mu_{S(difficult)} - zHR_{difficult}\right) \\ zFAR_{easy} &= \mu_{N(easy)} - C_{easy} = \mu_{N(easy)} - f\left(C_{difficult}\right) \\ &= \mu_{N(easy)} - f\left(\mu_{N(difficult)} - zFAR_{difficult}\right) \end{aligned}$$

In previous empirical experiments, the zHR$_{\text{easy}}$/zHR$_{\text{difficult}}$ and zFAR$_{\text{easy}}$/zFAR$_{\text{difficult}}$ plots are (at least approximately) linear (Glanzer et al., 2009, 2019; Hilford et al., 2015). In order to meet this requirement, $f(x)$ should be a linear function:

$$f(x) = bx + a$$

When $b \neq 1$, $f(x)$ can be rewritten as:

$$f(x) = x - (1 - b)\left(x - \frac{a}{1 - b}\right)$$

When $b < 1$, the decision criteria in the easy condition contract toward the point $a/(1\text{-}b)$ on the strength axis (compared with the difficult condition). When $b > 1$, the evidence strength criteria in the easy condition expand away from $a/(1\text{-}b)$. When $b = 1$, there is neither contraction nor expansion of evidence strength criteria in the easy condition. Instead, the evidence strength criteria simply shift by a distance of a.

The intercept of the zHR$_{\text{easy}}$/zHR$_{\text{difficult}}$ plot (denoted by h_{zHR}) is equal to the value of zHR$_{\text{easy}}$ when zHR$_{\text{difficult}}$ is 0:

$$h_{zHR} = \mu_{S(easy)} - f\left(\mu_{S(difficult)}\right) = \mu_{S(easy)} - b\mu_{S(difficult)} - a$$

Similarly, the intercept of the zFAR$_{\text{easy}}$/zFAR$_{\text{difficult}}$ plot (h_{zFAR}) is:

$$h_{zFAR} = \mu_{N(easy)} - f\left(\mu_{N(difficult)}\right) = \mu_{N(easy)} - b\mu_{N(difficult)} - a$$

The mirror effect holds when $h_{zHR} > 0$ and $h_{zFAR} < 0$, which is mathematically equivalent to the following inequality:

$$\mu_{N(easy)} - b\mu_{N(difficult)} < a < \mu_{S(easy)} - b\mu_{S(difficult)}$$

From this inequality, we can conclude that:

$$\mu_{N(easy)} - b\mu_{N(difficult)} < \mu_{S(easy)} - b\mu_{S(difficult)}$$

$$b < \frac{\mu_{S(easy)} - \mu_{N(easy)}}{\mu_{S(difficult)} - \mu_{N(difficult)}}$$

The distance between the means of signal and noise distributions should be longer in easy than difficult condition, and thus the maximum value of *b* is higher than 1. That is, the value of *b* can be lower than, higher than or equal to 1 to meet the requirement of the mirror effect, suggesting that the evidence strength criteria can either contract, expand or simply shift in the easy condition.

Figure S1 shows several examples of the mirror effect in equal variance SDT with different values of *b*. Here we set $\mu_{N(easy)} = \mu_{N(difficult)} = 0$, $\mu_{S(easy)} = 2$, $\mu_{S(difficult)} = 1$, and five evidence strength criteria in the difficult condition ($C_{difficult}$) as -0.75, 0, 0.75, 1.5 and 2.25. The value of *b* is 0.8 in Figure S1A, 1 in Figure S1B, and 1.2 in Figure S1C. Furthermore, *a* is equal to 0.5 in Figures S1A-S1C. The intercept is higher than 0 for the $zHR_{easy}/zHR_{difficult}$ plot and lower than 0 for the $zFAR_{easy}/zFAR_{difficult}$ plot in all of the three cases, revealing the mirror effect. However, the variance effect and zROC length effect are only observed in Figure S1A but not in Figures S1B-S1C. Specifically, the slope for the $zHR_{easy}/zHR_{difficult}$ and $zFAR_{easy}/zFAR_{difficult}$ plots is smaller than 1 in Figure S1A, equal to 1 in Figure S1B, and higher than 1 in Figure S1C. In addition, the zROC length in the easy condition is shorter than, equal to, and longer than that in the difficult condition in Figures S1A-S1C, respectively.

S2. Relationship between the three regularities and the change of log likelihood ratio (LR) criteria in equal variance SDT

In this section, we explain how to adjust the decision criteria on the objective log LR (λ) axis to obtain the three regularities in equal variance SDT. As discussed in the main text, the three regularities hold when the evidence strength criteria contract in easy (compared with difficult) condition. Here we demonstrate the relationship between the change of decision criteria on the strength axis and λ axis.

The probability density function for the signal (S) and noise (N) distributions can be written as (assuming both distributions have a variance of 1):

$$f(x|S) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_S)^2}{2}}$$
$$f(x|N) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_N)^2}{2}}$$

in which x represents evidence strength, and μ_S and μ_N are the mean of the signal and noise distribution, respectively. Then λ is calculated as:

$$\lambda = \log \frac{f(x|S)}{f(x|N)} = \frac{(x-\mu_N)^2}{2} - \frac{(x-\mu_S)^2}{2} = (\mu_S - \mu_N)x - \frac{{\mu_S}^2 - {\mu_N}^2}{2}$$

The relationship between x and λ can be rewritten as:

$$x = \frac{\lambda}{\mu_S - \mu_N} + \frac{\mu_S + \mu_N}{2}$$

Thus, the distance between decision criteria on the strength axis (denoted by C_{dist}) and the distance between criteria on the λ axis (denoted by λ_{dist}) have the following relationship:

$$C_{dist} = \frac{\lambda_{dist}}{\mu_S - \mu_N}$$

If the evidence decision criteria contract in the easy (compared with difficult) condition (i.e., $C_{dist(easy)} < C_{dist(difficult)}$), then the following inequality holds:

$$\frac{\lambda_{dist(easy)}}{\lambda_{dist(difficult)}} < \frac{\mu_{S(easy)} - \mu_{N(easy)}}{\mu_{S(difficult)} - \mu_{N(difficult)}}$$

The distance between the means of signal and noise distributions should be longer in easy than difficult condition, and thus the maximum value of $\lambda_{dist(easy)}/\lambda_{dist(difficult)}$ is higher than 1. That is, $\lambda_{dist(easy)}/\lambda_{dist(difficult)}$ can be lower than, higher than or equal to 1, suggesting that the decision criteria on the λ axis can either contract, expand or remain constant in the easy condition in order to obtain the three regularities.

S3. Fitting equal variance SDT to confidence rating data

In this section, we discuss how to compute the likelihood function for the equal variance SDT model fitted to confidence rating data in Experiments 1 and 2. Here we define the trials in which the correct stimulus is presented on the left side of the screen as “noise”, and the trials in which the correct stimulus is presented on the right side of the screen as “signal”. According to Section S2, the relationship between the decision criteria on the axis of evidence strength (denoted by *C*), and the criteria on the axis of log LR of the signal (S) and noise (N) distributions (denoted by λ), is:

$$C = \frac{\lambda}{\mu_S - \mu_N} + \frac{\mu_S + \mu_N}{2}$$

in which μ_S and μ_N are the mean of the signal and noise distributions, respectively. The probability of giving response using each point on the confidence rating scale when a trial comes from the signal or noise category can be computed as:

$$P_{conf=1,\,N} = P(\lambda < \lambda_1|N) = P(x < C_1|N) = \Phi(C_1 - \mu_N)$$

$$= \Phi\left(\frac{\lambda_1}{\mu_S - \mu_N} + \frac{\mu_S - \mu_N}{2}\right) = \Phi\left(\frac{\lambda_1}{d'} + \frac{d'}{2}\right)$$

$$P_{conf=1,\,S} = P(\lambda < \lambda_1|S) = P(x < C_1|S) = \Phi(C_1 - \mu_S) = \Phi\left(\frac{\lambda_1}{\mu_S - \mu_N} - \frac{\mu_S - \mu_N}{2}\right)$$

$$= \Phi\left(\frac{\lambda_1}{d'} - \frac{d'}{2}\right)$$

$$P_{conf=i(1<i<n),\,N} = P(\lambda_{i-1} < \lambda < \lambda_i|N) = P(C_{i-1} < x < C_i|N)$$

$$= \Phi(C_i - \mu_N) - \Phi(C_{i-1} - \mu_N) = \Phi\left(\frac{\lambda_i}{d'} + \frac{d'}{2}\right) - \Phi\left(\frac{\lambda_{i-1}}{d'} + \frac{d'}{2}\right)$$

$$P_{conf=i(1<i<n),\,S} = P(\lambda_{i-1} < \lambda < \lambda_i|S) = P(C_{i-1} < x < C_i|S)$$

$$= \Phi(C_i - \mu_S) - \Phi(C_{i-1} - \mu_S) = \Phi\left(\frac{\lambda_i}{d'} - \frac{d'}{2}\right) - \Phi\left(\frac{\lambda_{i-1}}{d'} - \frac{d'}{2}\right)$$

$$P_{conf=n,\,N} = P(\lambda > \lambda_{n-1}|N) = P(x > C_{n-1}|N) = \Phi(\mu_N - C_{n-1})$$

$$= \Phi\left(-\frac{\lambda_{n-1}}{d'} - \frac{d'}{2}\right)$$

$$P_{conf=n,\,S} = P(\lambda > \lambda_{n-1}|S) = P(x > C_{n-1}|S) = \Phi(\mu_S - C_{n-1})$$

$$= \Phi\left(-\frac{\lambda_{n-1}}{d'} + \frac{d'}{2}\right)$$

in which *x* represents the evidence strength, and *d'* is the distance between the means of the signal and noise distributions. The *n* refers to the total number of points on the confidence scale (6 in the current study). The likelihood of observing the confidence rating data can then be calculated as:

$$L = \prod_{i=1}^{n} P_{conf=i,\,N}{}^{m_{conf=i,\,N}} P_{conf=i,\,S}{}^{m_{conf=i,\,S}}$$

in which *m* represents the number of trials in each confidence category. The log likelihood function is:

$$l = \log L = \sum_{i=1}^{n} m_{conf=i,\,N} \log P_{conf=i,\,N} + \sum_{i=1}^{n} m_{conf=i,\,S} \log P_{conf=i,\,S}$$

We can find the parameter values that maximize the log likelihood function.

S4. Comparing the parameter-estimation and model-comparison approach

In this section, we used data simulation to compare the two methods to explore whether people set decision criteria based on objective LR, including the parameter-estimation approach and the model-comparison approach. We first simulated data from hypothetical experiments in which the objective LR corresponding to decision criteria differed across levels of task difficulty. Specifically, we simulated data from 100 hypothetical experiments, and in each experiment there were 50 participants completing two different signal detection tasks which could be characterized by equal variance SDT model. Each task contained 100 trials, and the two tasks differed in the level of difficulty (easy vs. difficult). Participants gave their response in the tasks using a 6-point confidence scale, as in our Experiments 1-3. The true parameters in the model were set based on the estimated parameter values in Experiment 1. Specifically, the value of *d'* was set as 2.23 in easy condition and 1.12 in difficult condition. Furthermore, the five criteria on the objective log LR (λ) axis were set as -3.07, -1.08, 0.22, 1.63 and 3.58 in easy condition, and -1.94, -0.74, 0, 0.74 and 1.97 in difficult condition. Thus, the objective λ criteria differed between the two conditions in the true model.

After simulating data from each hypothetical experiment, we examined the difference in the objective λ criteria between easy and difficult conditions using both the parameter-estimation and model-comparison approaches. For the parameter-estimation approach, we fit the equal variance SDT model separately to data from easy and difficult condition for each participant, and extracted the estimated value of the five λ criteria in each condition. Then we conducted 2 (condition: easy vs. difficult) × 5 (position: 1, 2, 3, 4 and 5) repeated measures ANOVA on the five λ criteria in each experiment. We were mostly interested in whether the condition × position interaction effect was statistically significant (after Greenhouse-Geisser correction).

For the model-comparison approach, we fit two different models to data from each participant. Model 1 assumed that decision criteria on the strength axis could be freely estimated in both easy and difficult conditions. Model 2 assumed that the evidence strength criteria across the two conditions were constrained such that they had the same objective LR. To compare the fit of the two models, for each participant we computed both the Akaike information criterion (AIC) and Bayesian information criterion (BIC) for each model, and then summed the AIC and BIC from all participants in each experiment separately for each model. Lower value of AIC and BIC indicates that the model could better fit the data.

Results revealed that the condition × position interaction effect on the five λ criteria reached statistical significance in all of the 100 hypothetical experiments, indicating that the parameter-estimation approach could effectively detect the difference between the objective λ criteria in easy and difficult conditions. However, the overall AIC and BIC for 50 participants were lower for Model 2 than Model 1 in all of the 100 hypothetical experiments. Thus, the model-comparison approach suggested that participants' decisions relied on the same objective λ criteria in easy and difficult conditions even when the difference in λ criteria across conditions existed in the true model. Furthermore, in the 5,000 participants from 100 experiments, 1,581 participants (less than 32%) showed higher AIC for Model 2 than Model 1, and only 13 participants (less than 1%) showed higher BIC for Model 2.

Next, we simulated data from 100 hypothetical experiments in which the objective λ criteria remained constant across difficulty levels. Specifically, each of the five criteria on the objective λ axis for both easy and difficult conditions in the true model was set as the mean of the estimated parameter values in the two conditions of Experiment 1. For each simulated experiment, we used both the parameter-estimation and model-comparison approaches to compare the objective λ criteria in the two conditions. Results from the parameter-estimation approach revealed that the condition × position interaction effect on the five λ criteria reached statistical significance in only 2 of the 100 hypothetical experiments, indicating an acceptable Type I error rate (0.02). For the model-comparison approach, the overall AIC and BIC for 50 participants were lower for Model 2 than Model 1 in all hypothetical experiments. Furthermore, in the 5,000 participants from 100 experiments, only 315 participants showed higher AIC for Model 2 than Model 1, and no participant showed higher BIC for Model 2. Thus, both approaches could detect the null difference in objective λ criteria between conditions.

Based on the results above, we could conclude that the parameter-estimation approach could accurately detect whether there was reliable difference in objective λ criteria across experimental conditions. In contrast, the model-comparison approach lacked statistical power, and preferred the model with constant λ criteria across conditions regardless of whether there was indeed null difference in the true model. Thus, we used the parameter-estimation approach in our Experiments 1-3.

S5. Fitting unequal variance SDT model to data in Experiment 3

In this section, we introduce how to fit unequal variance SDT model with decision criteria on the objective LR axis to confidence rating data in Experiment 3, and then compute the parametric measure of the area under the ROC curve (A_z) using the estimated parameter values. For mathematical simplicity, here we assume the signal (S) distribution has a mean of *d'* and a standard deviation of σ_S, and the noise (N) distribution has a mean of 0 and a standard deviation of 1 (Glanzer et al., 2009). Then the probability density function for the signal and noise distributions can be written as:

$$f(x|S) = \frac{1}{\sqrt{2\pi}\sigma_S} e^{-\frac{(x-d')^2}{2\sigma_S{}^2}}$$

$$f(x|N) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

in which *x* represents evidence strength. Then the objective log LR (λ) for each value of *x* is calculated as (see also Glanzer et al., 2009):

$$\lambda(x) = \log \frac{f(x|S)}{f(x|N)} = -\log \sigma_S + \frac{1}{2}\left[x^2 - \frac{(x-d')^2}{\sigma_S{}^2}\right]$$

$$= \frac{(\sigma_S{}^2 - 1)}{2\sigma_S{}^2}x^2 + \frac{d'}{\sigma_S{}^2}x - \frac{d'^2}{2\sigma_S{}^2} - \log \sigma_S$$

The $\lambda(x)$ is a quadratic function. When $\sigma_S > 1$, $\lambda(x)$ has a minimum value λ^* at $x = C^* = - d' / (\sigma_S^2 - 1)$. When $\sigma_S < 1$, $\lambda(x)$ has a maximum value λ^* at $x = C^*$. The value of λ^* is equal to:

$$\lambda^* = \lambda(C^*) = \frac{d'^2}{2(1-\sigma_S{}^2)} - \log \sigma_S$$

Thus, the decision criterion on the λ axis is mapped to only one criterion (C^*) on the *x* axis when $\lambda = \lambda^*$. When $\lambda \neq \lambda^*$, each λ criterion corresponds to two different criteria on the *x* axis, and the distance between C^* and each of these two evidence strength criteria is the same (denoted by C_{dist}). Thus, the five λ criteria for a 6-point confidence rating scale (λ_1, λ_2, λ_3, λ_4 and λ_5) are mathematically equivalent to five different values of C_{dist} (C_{dist1}, C_{dist2}, C_{dist3}, C_{dist4} and C_{dist5}).

When $\sigma_S > 1$, people judge a stimulus as noise when x is between the two evidence strength criteria corresponding to the same λ criterion, and as signal when x is outside these two strength criteria. Thus, the probability of giving response using each point on the confidence rating scale when a stimulus comes from the signal or noise category can be computed as:

$$P_{conf=1,\,N} = P(\lambda < \lambda_1|N) = P(C^* - C_{dist1} < x < C^* + C_{dist1}|N)$$

$$= \Phi(C^* + C_{dist1}) - \Phi(C^* - C_{dist1})$$

$$P_{conf=1,\,S} = P(\lambda < \lambda_1|S) = P(C^* - C_{dist1} < x < C^* + C_{dist1}|S)$$

$$= \Phi\left(\frac{C^* + C_{dist1} - d'}{\sigma_S}\right) - \Phi\left(\frac{C^* - C_{dist1} - d'}{\sigma_S}\right)$$

$$P_{conf=i(1<i<n),\,N} = P(\lambda_{i-1} < \lambda < \lambda_i|N)$$

$$= P\big(C^* - C_{dist(i)} < x < C^* - C_{dist(i-1)}\big|N\big)$$

$$+ P\big(C^* + C_{dist(i-1)} < x < C^* + C_{dist(i)}\big|N\big)$$

$$= \Phi\big(C^* - C_{dist(i-1)}\big) - \Phi\big(C^* - C_{dist(i)}\big) + \Phi\big(C^* + C_{dist(i)}\big)$$

$$- \Phi\big(C^* + C_{dist(i-1)}\big)$$

$$P_{conf=i(1<i<n),\,S} = P(\lambda_{i-1} < \lambda < \lambda_i|S)$$

$$= P\big(C^* - C_{dist(i)} < x < C^* - C_{dist(i-1)}\big|S\big)$$

$$+ P\big(C^* + C_{dist(i-1)} < x < C^* + C_{dist(i)}\big|S\big)$$

$$= \Phi\left(\frac{C^* - C_{dist(i-1)} - d'}{\sigma_S}\right) - \Phi\left(\frac{C^* - C_{dist(i)} - d'}{\sigma_S}\right)$$

$$+ \Phi\left(\frac{C^* + C_{dist(i)} - d'}{\sigma_S}\right) - \Phi\left(\frac{C^* + C_{dist(i-1)} - d'}{\sigma_S}\right)$$

$$\begin{aligned} P_{conf=n,\,N} &= P(\lambda > \lambda_{n-1}|N) \\ &= P\big(x < C^* - C_{dist(n-1)}\big|N\big) + P\big(x > C^* + C_{dist(n-1)}\big|N\big) \\ &= \Phi\big(C^* - C_{dist(n-1)}\big) + 1 - \Phi\big(C^* + C_{dist(n-1)}\big) \end{aligned}$$

$$\begin{aligned} P_{conf=n,\,S} &= P(\lambda > \lambda_{n-1}|S) \\ &= P\big(x < C^* - C_{dist(n-1)}\big|S\big) + P\big(x > C^* + C_{dist(n-1)}\big|S\big) \\ &= \Phi\left(\frac{C^* - C_{dist(n-1)} - d'}{\sigma_S}\right) + 1 - \Phi\left(\frac{C^* + C_{dist(n-1)} - d'}{\sigma_S}\right) \end{aligned}$$

in which *n* refers to the total number of points on the confidence scale (6 in the current study).

When $\sigma_S < 1$, people judge a stimulus as signal when x is between the two evidence strength criteria corresponding to the same λ criterion, and as noise when x is outside these two strength criteria. Thus, the probability of giving response using each point on the confidence rating scale when a stimulus is signal or noise can be computed as:

$$\begin{aligned} P_{conf=1,\,N} &= P(\lambda < \lambda_1|N) = P(x < C^* - C_{dist1}|N) + P(x > C^* + C_{dist1}|N) = \\ &= \Phi(C^* - C_{dist1}) + 1 - \Phi(C^* + C_{dist1}) \end{aligned}$$

$$\begin{aligned} P_{conf=1,\,S} &= P(\lambda < \lambda_1|S) = P(x < C^* - C_{dist1}|S) + P(x > C^* + C_{dist1}|S) = \\ &= \Phi\left(\frac{C^* - C_{dist1} - d'}{\sigma_S}\right) + 1 - \Phi\left(\frac{C^* + C_{dist1} - d'}{\sigma_S}\right) \end{aligned}$$

$$\begin{aligned} P_{conf=i(1<i<n),\,N} &= P(\lambda_{i-1} < \lambda < \lambda_i|N) \\ &= P\big(C^* - C_{dist(i-1)} < x < C^* - C_{dist(i)}\big|N\big) \\ &+ P\big(C^* + C_{dist(i)} < x < C^* + C_{dist(i-1)}\big|N\big) \end{aligned}$$

$$= \Phi\left(C^* - C_{dist(i)}\right) - \Phi\left(C^* - C_{dist(i-1)}\right) + \Phi\left(C^* + C_{dist(i-1)}\right)$$

$$- \Phi\left(C^* + C_{dist(i)}\right)$$

$$P_{conf=i(1<i<n),\,S} = P(\lambda_{i-1} < \lambda < \lambda_i|S)$$

$$= P\left(C^* - C_{dist(i-1)} < x < C^* - C_{dist(i)}\middle|S\right)$$

$$+ P\left(C^* + C_{dist(i)} < x < C^* + C_{dist(i-1)}\middle|S\right)$$

$$= \Phi\left(\frac{C^* - C_{dist(i)} - d'}{\sigma_S}\right) - \Phi\left(\frac{C^* - C_{dist(i-1)} - d'}{\sigma_S}\right)$$

$$+ \Phi\left(\frac{C^* + C_{dist(i-1)} - d'}{\sigma_S}\right) - \Phi\left(\frac{C^* + C_{dist(i)} - d'}{\sigma_S}\right)$$

$$P_{conf=n,\,N} = P(\lambda > \lambda_{n-1}|N) = P\left(C^* - C_{dist(n-1)} < x < C^* + C_{dist(n-1)}\middle|N\right)$$

$$= \Phi\left(C^* + C_{dist(n-1)}\right) - \Phi\left(C^* - C_{dist(n-1)}\right)$$

$$P_{conf=n,\,S} = P(\lambda > \lambda_{n-1}|S) = P\left(C^* - C_{dist(n-1)} < x < C^* + C_{dist(n-1)}\middle|S\right)$$

$$= \Phi\left(\frac{C^* + C_{dist(n-1)} - d'}{\sigma_S}\right) - \Phi\left(\frac{C^* - C_{dist(n-1)} - d'}{\sigma_S}\right)$$

The likelihood of observing the confidence rating data can then be calculated as:

$$L = \prod_{i=1}^{n} P_{conf=i,\,N}{}^{m_{conf=i,\,N}} P_{conf=i,\,S}{}^{m_{conf=i,\,S}}$$

in which *m* represents the number of trials in each confidence category. The log likelihood function is:

$$l = \log L = \sum_{i=1}^{n} m_{conf=i,\,N} \log P_{conf=i,\,N} + \sum_{i=1}^{n} m_{conf=i,\,S} \log P_{conf=i,\,S}$$

We can find the parameter values that maximize the log likelihood function. After getting the estimated values of the five C_{dist}, we can then obtain the location of the two evidence strength criteria corresponding to each λ criterion, and compute the value of the five λ criteria using the equation for $\lambda(x)$ described above.

Next, we introduce how to approximate the value of A_z using trapezoidal rule based on estimated parameter values. When the estimated value of σ_S is higher than 1, we generate 50,000 equal-interval values of λ between λ^* and 50. When the estimated value of σ_S is lower than 1, we generate 50,000 equal-interval values of λ between -50 and λ^*. For each value of λ, we computed the two corresponding evidence strength criteria, and the value of HR and FAR. We then connect the neighboring HR/FAR points on the ROC curve, as well as the (0, 0) and (1, 1) points, with straight lines. Then A_z is approximated as the area under the straight lines.

S6. An example of the Bayesian weighting model

In this section, we introduce an example of the Bayesian weighting model in equal variance SDT. Here we assume that the observed evidence strength x_{obs} comes from a normal distribution with a mean of the true evidence strength x_{true} and a standard deviation of σ_x. According to Section S2, the relationship between evidence strength x and the objective log LR λ in equal variance SDT can be written as:

$$x = \frac{\lambda}{\mu_S - \mu_N} + \frac{\mu_S + \mu_N}{2} = \frac{\lambda}{d'} + \frac{\mu_S + \mu_N}{2}$$

in which μ_S and μ_N are the mean of the signal and noise distribution, respectively, and d' is the distance between the two distributions. Thus, the observed log LR λ_{obs} is also distributed as a normal distribution with a mean of the true log LR λ_{true} and a standard deviation of σ_λ:

$$\lambda_{obs} \sim N(\lambda_{true}, {\sigma_\lambda}^2)$$

$$\sigma_\lambda = \sigma_x \cdot d'$$

Furthermore, we assume that people's prior belief about λ_{true} is a normal distribution with a mean of λ_{prior} and a standard deviation of σ_{prior}:

$$\lambda_{true} \sim N\left(\lambda_{prior}, {\sigma_{prior}}^2\right)$$

According to the Bayesian weighting model, people infer the posterior distribution of λ_{true} by combining the observed log LR and their prior belief through a Bayesian inference process. Because the normal prior distribution is the conjugate prior of a normal distribution (Murphy, 2007), we can directly compute the mean of the posterior distribution λ_{post}:

$$\lambda_{post} = E(\lambda_{true}|\lambda_{obs}) = \frac{1/{\sigma_\lambda}^2}{1/{\sigma_\lambda}^2 + 1/{\sigma_{prior}}^2}\lambda_{obs} + \frac{1/{\sigma_{prior}}^2}{1/{\sigma_\lambda}^2 + 1/{\sigma_{prior}}^2}\lambda_{prior}$$

Thus, λ_{post} is a linear function of the observed log LR λ_{obs} (as suggested by the LLO function), and the slope γ of the LLO function is:

$$\gamma = \frac{1/{\sigma_\lambda}^2}{1/{\sigma_\lambda}^2 + 1/{\sigma_{prior}}^2} = \frac{{\sigma_{prior}}^2}{{\sigma_{prior}}^2 + {\sigma_\lambda}^2} = \frac{{\sigma_{prior}}^2}{{\sigma_{prior}}^2 + {\sigma_x}^2 d'^2}$$

The equation above indicates that the value of γ decreases when *d'* increases (i.e., when the task difficulty decreases).

S7. Comparing the evidence strength criteria in easy and difficult conditions

In this section, we quantitatively compared the decision criteria on the axis of evidence strength in easy and difficult conditions separately in Experiments 1-3. Data in the first two experiments could be characterized by the equal variance SDT model in 2AFC tasks. To estimate the evidence strength criteria, here we defined the means of the signal and noise distributions as *d'*/2 and -*d'*/2, respectively. Then the location of the five evidence strength criteria C could be computed based on the estimated objective log LR criteria λ (see Section S2):

$$C = \frac{\lambda}{d'}$$

After computing the value of C separately in each experimental condition for each participant, we performed 2 (condition: easy vs. difficult) × 5 (position: 1, 2, 3, 4 and 5) repeated measures ANOVA on C in Experiments 1 and 2. Results revealed that the main effect of position was statistically significant in both experiments, Fs > 174.26, ps < .001, $\eta_p^2 > .78$, $\text{BF}_{\text{incl}} > 100$, indicating that the value of C increased along the axis from left to right. Furthermore, the main effect of condition also reached statistical significance in Experiment 1, $F(1, 49) = 11.18$, $p = .002$, $\eta_p^2 = .19$, $\text{BF}_{\text{incl}} = 3.07$, suggesting the overall value of C was higher in easy than difficult condition. However, there was no reliably main effect of condition in Experiment 2, $F(1, 49) = 0.61$, $p = .437$, $\eta_p^2 = .01$, $\text{BF}_{\text{incl}} = 0.16$. More importantly, the interaction effect between position and condition was close to significance in Experiment 1, $F(1.40, 68.35) = 3.04$, $p = .072$, $\eta_p^2 = .06$, $\text{BF}_{\text{incl}} = 2.93$ (Greenhouse-Geisser corrected), and statistically detectable in Experiment 2, $F(1.69, 82.80) = 10.17$, $p < .001$, $\eta_p^2 = .17$, $\text{BF}_{\text{incl}} > 100$ (Greenhouse-Geisser corrected).

Further simple effect analyses were then conducted separately for Experiments 1 and 2. In Experiment 1, C was significantly lower in difficult than easy condition at Positions 1-3 based on p value (although the Bayes factor for Position 2 was inconclusive), ts > 2.03, ps < .05, Cohen's $d > 0.28$; $\text{BF}_{10} > 6.70$ for Positions 1 and 3, and $\text{BF}_{10} = 1.01$ for Position 2. In addition, the value of C did not differ between conditions at Positions 4-5, *t*s < 0.96, *p*s > .34, Cohen's *d* < 0.14, BF_{10} < 0.24 (see Figure S2). In Experiment 2, *C* was significantly lower in difficult than easy condition at Positions 1-2 based on *p* value (although the Bayes factor for Position 2 was inconclusive), *t*s > 2.52, *p*s < .05, Cohen's *d* > 0.35; BF_{10} = 31.02 for Positions 1, and BF_{10} = 2.68 for Position 2. At Positions 3 and 4, *C* did not reliably differ between conditions, *t*s < 2.01, *p*s > .05, Cohen's *d* < 0.29, BF_{10} < 0.97. Furthermore, C was higher in difficult than easy condition at Position 5, *t*(49) = 2.97, *p* = .005, Cohen's *d* = 0.42, BF_{10} = 7.40 (see Figure S3). In summary, the evidence strength criteria tended to fan out in difficult (compared with easy) condition in Experiments 1 and 2.

Data in Experiment 3 were fitted by an unequal variance SDT model which assumed that participants set decision criteria on the axis of evidence strength (rather than objective LR). The model assumed that the noise distribution had a mean of 0 and a standard deviation of 1, and the signal distribution had a mean of *d'* and a standard deviation of σ_S (Glanzer et al., 2009). We fit the model to confidence rating data separately in each experimental condition for each participant, and estimated the five evidence strength criteria *C*. Then we performed 2 (condition: easy vs. difficult) × 5 (position: 1, 2, 3, 4 and 5) repeated measures ANOVA on *C*, which revealed that the main effects of both position and condition were statistically detectable, *F*s > 20.31, *p*s < .001, η_p^2 > .20, $\mathrm{BF}_{\mathrm{incl}}$ > 100. More importantly, the interaction effect between position and condition was also reliable, *F*(1.99, 97.60) = 8.62, *p* < .001, η_p^2 = .15, $\mathrm{BF}_{\mathrm{incl}}$ > 100 (Greenhouse-Geisser corrected). Further analyses showed that *C* was significantly lower in difficult than easy condition at Positions 1-4 based on *p* value (although the Bayes factor for Position 4 was inconclusive), *t*s > 2.40, *p*s < .05, Cohen's *d* > 0.34; BF_{10} > 78.02 for Positions 1-3, and BF_{10} = 2.11 for Position 4. Furthermore, *C* did not reliably differ between conditions at Position 5, *t*(49) = 0.54, *p* = . 591, Cohen's *d* = 0.08, BF_{10} = 0.18 (see Figure S4). Thus, the evidence strength criteria also fanned out in difficult condition in Experiment 3.

For completeness, we also compared the objective log LR (λ) corresponding to the five estimated evidence strength criteria *C* across experimental conditions in Experiment 3. Results from the 2 (condition: easy vs. difficult) × 5 (position: 1, 2, 3, 4 and 5) repeated measures ANOVA indicated a significant main effect of position, *F*(1.29, 63.36) = 61.32, *p* < .001, η_p^2 = .56, BF_{incl} > 100 (Greenhouse-Geisser corrected), but a non-significant main effect of condition, *F*(1, 49) = 2.65, *p* = .110, η_p^2 = .05, BF_{incl} = 0.30. More importantly, there was a reliable interaction effect between condition and position, *F*(2.15, 105.27) = 6.99, *p* = .001, η_p^2 = .13, BF_{incl} > 100 (Greenhouse-Geisser corrected). Further analyses revealed that λ was reliably lower in easy than difficult condition at Positions 1 and 2, *t*s > 3.60, *p*s < .005, Cohen's *d* > 0.47, BF_{10} > 21.11. At Positions 3-5, λ was numerically higher in easy than difficult condition. However, this difference did not reach statistical significance at Positions 3 and 5, *t*s < 1.24, *p*s > .22, Cohen's *d* < 0.18, BF_{10} < 0.32. Furthermore, λ was significantly higher in easy condition at Position 4 based on *p* value (although the Bayes factor was inconclusive), *t*(49) = 2.44, *p* = .019, Cohen's *d* = 0.35, BF_{10} = 2.22 (see Figure S5). These results suggest that the objective log LR for decision criteria tended to fan out in easy (compared with difficult) condition, consistent with the findings in the main text.

References

Glanzer, M., Hilford, A., Kim, K., & Maloney, L. T. (2019). Generality of likelihood ratio decisions. *Cognition*, *191*, 103931. https://doi.org/https://doi.org/10.1016/j.cognition.2019.03.023

Glanzer, M., Hilford, A., & Maloney, L. T. (2009). Likelihood ratio decisions in memory: Three implied regularities. *Psychonomic Bulletin & Review*, *16*(3), 431–455. https://doi.org/10.3758/PBR.16.3.431

Hilford, A., Maloney, L. T., Glanzer, M., & Kim, K. (2015). Three regularities of recognition memory: the role of bias. *Psychonomic Bulletin & Review*, *22*(6), 1646–1664. https://doi.org/10.3758/s13423-015-0829-0

Murphy, K. P. (2007). *Conjugate Bayesian Analysis of the Gaussian Distribution*. https://www.cs.ubc.ca/~murphyk/Papers/bayesGauss.pdf

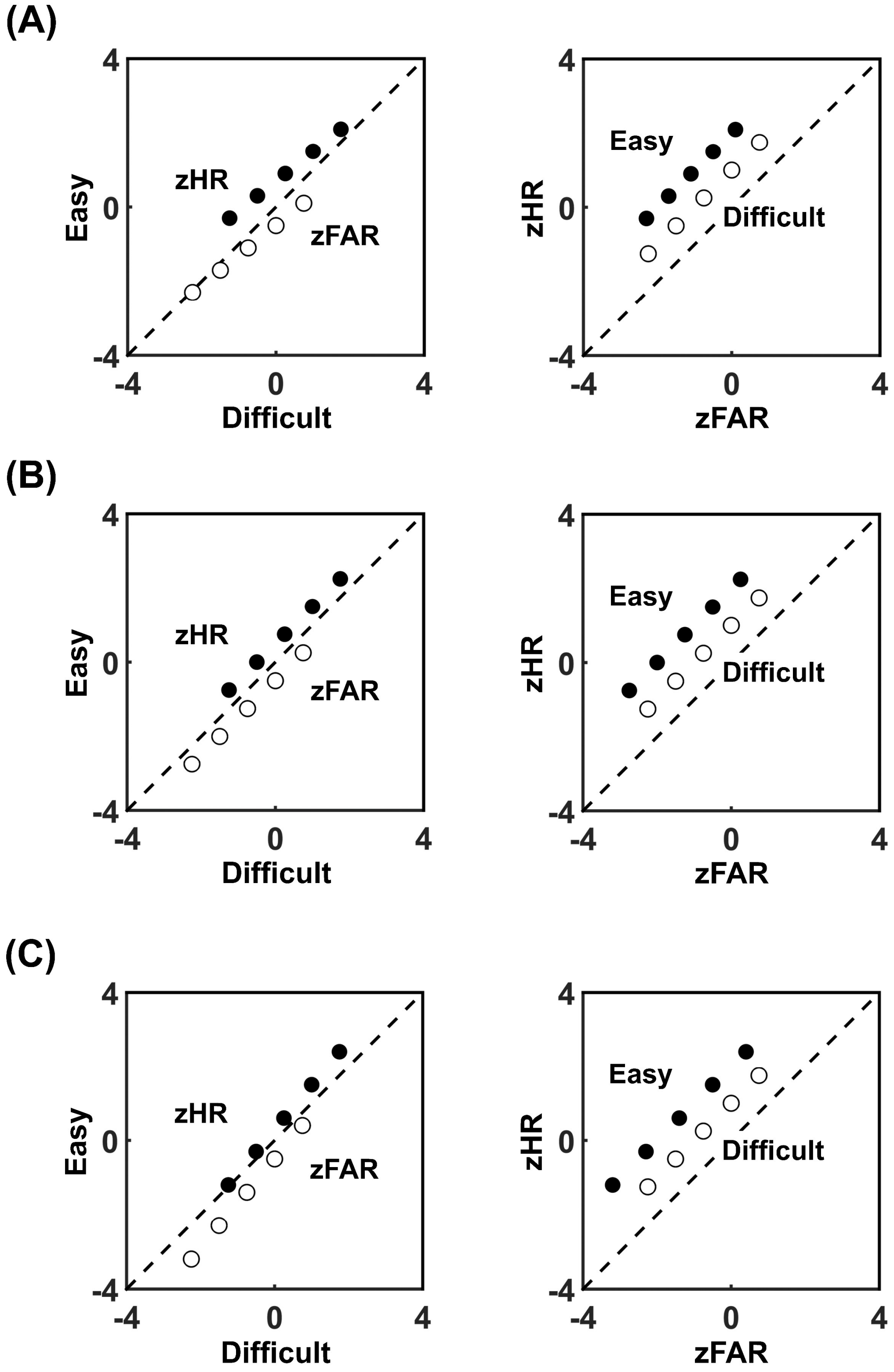

Figure S1. The $\text{zHR}_{\text{easy}}/\text{zHR}_{\text{difficult}}$ and $\text{zFAR}_{\text{easy}}/\text{zFAR}_{\text{difficult}}$ plots (left) and zHR/zFAR plots (right) when *b* is equal to 0.8 (A), 1 (B) or 1.2 (C). See Section S1 for details.

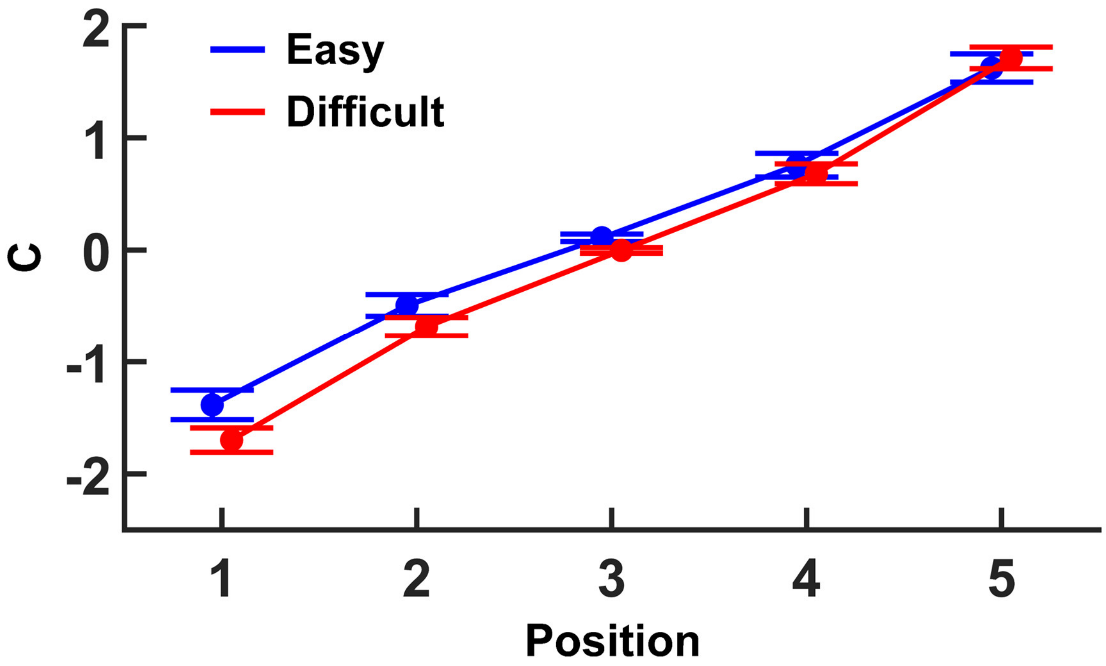

Figure S2. The evidence strength criteria *C* as a function of experimental condition (easy vs. difficult) and position (from 1 to 5) in Experiment 1. Error bars represent standard errors.

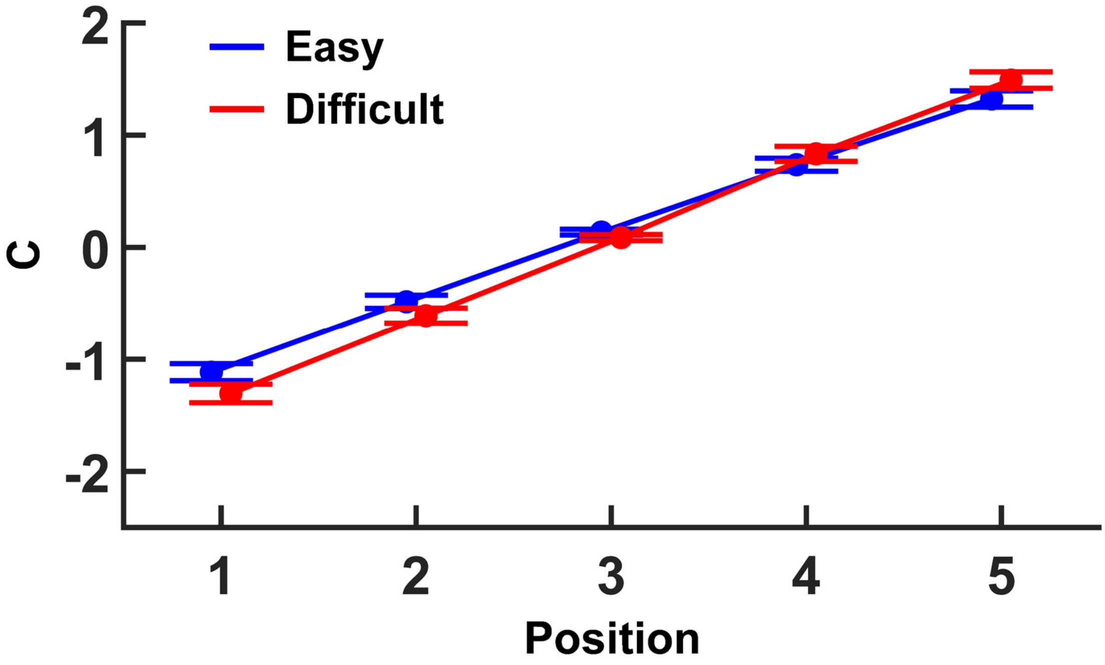

Figure S3. The evidence strength criteria *C* as a function of experimental condition (easy vs. difficult) and position (from 1 to 5) in Experiment 2. Error bars represent standard errors.

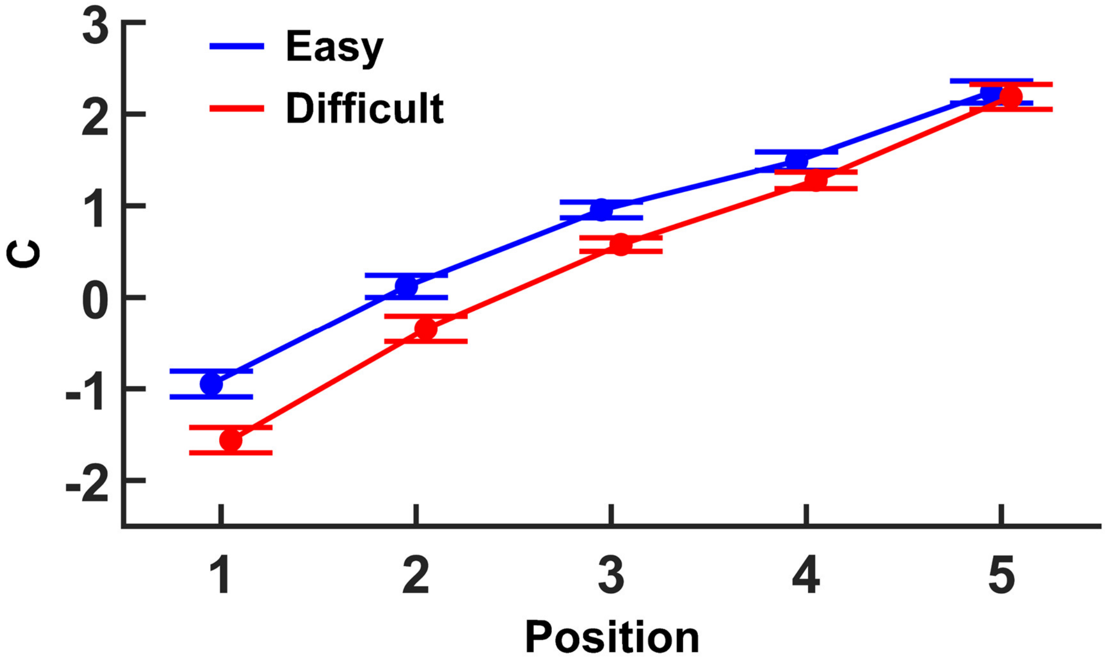

Figure S4. The evidence strength criteria *C* as a function of experimental condition (easy vs. difficult) and position (from 1 to 5) in Experiment 3. Error bars represent standard errors.

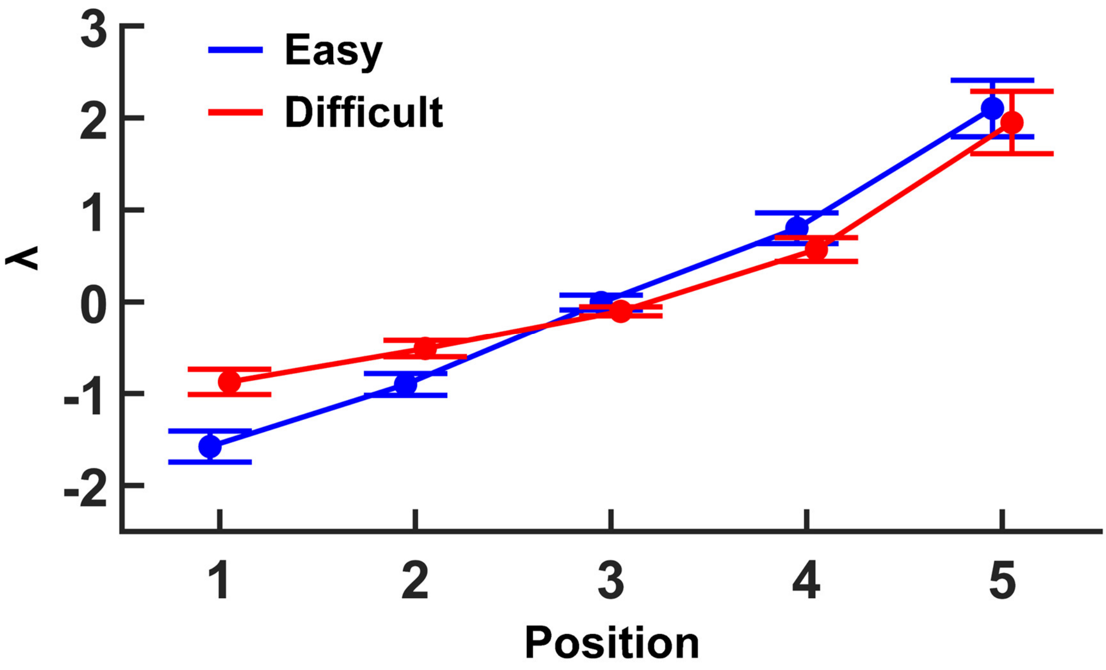

Figure S5. The objective log LR *λ* corresponding to the evidence strength criteria as a function of experimental condition (easy vs. difficult) and position (from 1 to 5) in Experiment 3. Error bars represent standard errors.