Supplementary Material for

Modelling Perceptual Confidence and the Confidence Forced-Choice Paradigm

Pascal Mamassian[1] and Vincent de Gardelle[2]

(1) Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France
(2) CNRS and Paris School of Economics, Paris, France

Appendix A: Properties of the Probability of being Self-Consistent

In this section, we present the derivations of the probability of being self-consistent and we relate it to the confidence probability that we defined in the text.

A.1. Probability of being Self-Consistent for some Sensory Evidence

What is the probability that the observer's perceptual decision is self-consistent, given some sensory evidence s? Here, self-consistent refers to the perceptual decision that matches the most frequent decision for a particular stimulus μ_s. In the framework of signal detection theory used here (see section 4.a), there is a single most frequent decision, or perceptual mode *M*, which is

$$M(\mu_s) = \begin{cases} \text{'R'} & \text{if } \mu_s > \theta_s \,, \\ \text{'L'} & \text{otherwise} \end{cases} . \tag{S1}$$

Therefore, the probability that the observer's perceptual decision is self-consistent, given that she has access to sensory evidence s is

$$\begin{aligned} \mathrm{P}(\text{self-consistent} \mid s) &= \textstyle\sum_{\mu_s} \left(\mathrm{P}(\text{self-consistent} \mid \mu_s, s) \;\; \mathrm{P}(\,\mu_s \mid s\,) \right) \\ &= \textstyle\sum_{\mu_s} \left(\mathrm{P}(D = M(\mu_s) \mid \mu_s, s) \;\; \mathrm{P}(\,\mu_s \mid s\,) \right) \\ &= \textstyle\sum_{\mu_s > \theta_s} \left(\mathrm{P}(D = \text{'R'} \mid s) \;\; \mathrm{P}(\,\mu_s \mid s\,) \right) + \sum_{\mu_s \le \theta_s} \left(\mathrm{P}(D = \text{'L'} \mid s) \;\; \mathrm{P}(\,\mu_s \mid s\,) \right) \\ &= \begin{cases} \sum_{\mu_s > \theta_s} \mathrm{P}(\,\mu_s \mid s\,) & \text{if } s > \theta_s \,, \\ \sum_{\mu_s \le \theta_s} \mathrm{P}(\,\mu_s \mid s\,) & \text{otherwise} \end{cases} \end{aligned} . \tag{S2}$$

Using Bayes' rule, we have

$$\mathrm{P}(\,\mu_s \mid s\,) = \mathrm{P}(s \mid \mu_s)\,\mathrm{P}(\mu_s) / \,\mathrm{P}(s) = \frac{\varphi(s;\,\mu_s,\sigma_s^2)\,\mathrm{P}(\mu_s)}{\sum_{\mu_s}\left(\varphi(s;\,\mu_s,\sigma_s^2)\,\mathrm{P}(\mu_s)\right)} \,, \tag{S3}$$

where $\varphi(x;\,\mu_s,\sigma_s^2)$ is the probability distribution function of the normal distribution with mean μ_s and variance σ_s^2, so that Equation S2 can be rewritten as

$$\mathrm{P}(\text{self-consistent} \mid s) = \begin{cases} \dfrac{\sum_{\mu_s > \theta_s}\left(\varphi(s;\,\mu_s,\sigma_s^2)\,\mathrm{P}(\mu_s)\right)}{\sum_{\mu_s}\left(\varphi(s;\,\mu_s,\sigma_s^2)\,\mathrm{P}(\mu_s)\right)} & \text{if } s > \theta_s \,, \\ \dfrac{\sum_{\mu_s \le \theta_s}\left(\varphi(s;\,\mu_s,\sigma_s^2)\,\mathrm{P}(\mu_s)\right)}{\sum_{\mu_s}\left(\varphi(s;\,\mu_s,\sigma_s^2)\,\mathrm{P}(\mu_s)\right)} & \text{otherwise} \end{cases} . \tag{S4}$$

Figure A1 shows the probability of being self-consistent as a function of sensory evidence when there are seven possible stimuli with varying strengths that can occur with equal probability (see parameters in Table 2).

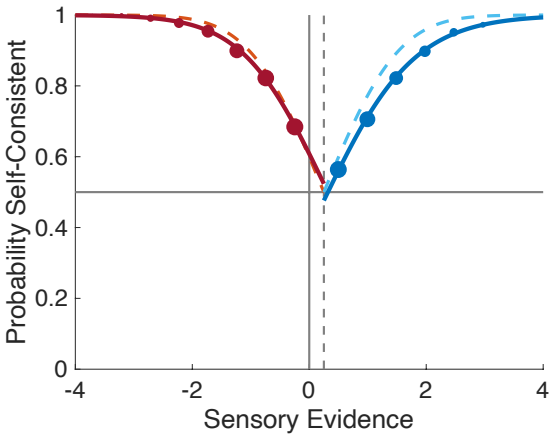

Figure A1. Relationship between self-consistency and sensory evidence. Dots represent 100,000 simulated trials where the perceptual outcome is either 'R' (blue dots) or 'L' (red dots), discretized in bins of size 0.5 units of sensory evidence. The dark blue and red continuous curves show the probability of being self-consistent (Equation S4) obtained in the conditions of the simulations (7 stimulus strengths). The light blue and red dashed curves show the limiting scenario of Equation S7 where any stimulus strength can be presented. In this plot, parameter values are those listed in Table 2.

There is one common special case where Equation S2 is simplified. When a stimulus can have any strength, and when all of these strengths have an equal probability of occurrence, then this equation becomes

$$\mathrm{P}(\text{self-consistent} \mid s) = \begin{cases} \Phi((s-\theta_s)/\sigma_s) & \text{if } s > \theta_s \\ 1-\Phi((s-\theta_s)/\sigma_s) & \text{otherwise} \end{cases}, \tag{S5}$$

where Φ is the cumulative of the standard normal distribution. In this equation, we recognize the confidence evidence for the ideal observer that we defined in Equation 3 of the main text, so that Equation S5 can be rewritten as

$$\mathrm{P}(\text{self-consistent} \mid s) = \begin{cases} \Phi(w_{\text{ideal}}) & \text{if } w_{\text{ideal}} > 0 \\ 1-\Phi(w_{\text{ideal}}) & \text{otherwise} \end{cases}, \tag{S6}$$

or in short

$$\mathrm{P}(\text{self-consistent} \mid s) \;=\; \Phi(|w_{\text{ideal}}|) \quad . \tag{S7}$$

If instead of self-consistency we were interested in accuracy, then we would have replaced the condition $(\mu_s > \theta_s)$ by $(\mu_s > 0)$ in Equation S1, so that

$$\mathrm{P}(\text{correct} \mid s) = \begin{cases} \Phi(s/\sigma_s) & \text{if } s > \theta_s\,, \\ 1 - \Phi(s/\sigma_s) & \text{otherwise} \end{cases} . \tag{S8}$$

A.2. Probability of being Self-Consistent for some Sensory Strength

If instead of focusing on one single trial where the sensory evidence was s, we were interested in the probability that the observer's perceptual decision is self-consistent for a given displayed stimulus μ_s, then

$$\begin{aligned} \mathrm{P}(\text{self-consistent} \mid \mu_s) &= \int_{-\infty}^{+\infty} \mathrm{P}(\text{self-consistent} \mid \mu_s, s)\, \mathrm{P}(s \mid \mu_s)\, ds \\ &= \int_{-\infty}^{+\infty} \mathrm{P}(D = M(\mu_s) \mid \mu_s, s)\, \mathrm{P}(s \mid \mu_s)\, ds \\ &= \begin{cases} \int_{-\infty}^{+\infty} \mathrm{P}(D = \text{'R'} \mid s)\, \mathrm{P}(s \mid \mu_s)\, ds & \text{if } \mu_s > \theta_s\,, \\ \int_{-\infty}^{+\infty} \mathrm{P}(D = \text{'L'} \mid s)\, \mathrm{P}(s \mid \mu_s)\, ds & \text{otherwise} \end{cases} \\ &= \begin{cases} \int_{\theta_s}^{+\infty} \varphi(s;\, \mu_s, \sigma_s^2)\, ds \;= \Phi((\mu_s - \theta_s)/\sigma_s) & \text{if } \mu_s > \theta_s\,, \\ \int_{-\infty}^{\theta_s} \varphi(s;\, \mu_s, \sigma_s^2)\, ds \;= 1 - \Phi((\mu_s - \theta_s)/\sigma_s) & \text{otherwise} \end{cases} \\ &= \Phi(|\mu_s - \theta_s|/\sigma_s) \end{aligned} \tag{S9}$$

A.3. Probability of being Self-Consistent for some Confidence Evidence

We consider next the probability of being self-consistent given the confidence evidence and the perceptual decision on the current trial. This probability of being self-consistent is

$$\mathrm{P}(\text{self-consistent} \mid w, D) = \frac{\mathrm{P}(\text{self-consistent}, w, D)}{\mathrm{P}(w, D)}$$

$$= \frac{\sum_{\mu_s}\left(\int_{-\infty}^{+\infty} \mathrm{P}(\text{self-consistent}, w, D \mid \mu_s, s)\, \mathrm{P}(s \mid \mu_s)\, \mathrm{P}(\mu_s)\, ds\right)}{\sum_{\mu_s}\left(\int_{-\infty}^{+\infty} \mathrm{P}(w, D \mid \mu_s, s)\, \mathrm{P}(s \mid \mu_s)\, \mathrm{P}(\mu_s)\, ds\right)}$$

$$= \begin{cases} \dfrac{\sum_{\mu_s > \theta_s}\left(\int_{\theta_s}^{+\infty} \mathrm{P}(w \mid \mu_s, s)\, \mathrm{P}(s \mid \mu_s)\, \mathrm{P}(\mu_s)\, ds\right)}{\sum_{\mu_s}\left(\int_{\theta_s}^{+\infty} \mathrm{P}(w \mid \mu_s, s)\, \mathrm{P}(s \mid \mu_s)\, \mathrm{P}(\mu_s)\, ds\right)} & \text{if } D = \text{'R'} \\ \dfrac{\sum_{\mu_s \le \theta_s}\left(\int_{-\infty}^{\theta_s} \mathrm{P}(w \mid \mu_s, s)\, \mathrm{P}(s \mid \mu_s)\, \mathrm{P}(\mu_s)\, ds\right)}{\sum_{\mu_s}\left(\int_{-\infty}^{\theta_s} \mathrm{P}(w \mid \mu_s, s)\, \mathrm{P}(s \mid \mu_s)\, \mathrm{P}(\mu_s)\, ds\right)} & \text{otherwise} \end{cases} . \quad \text{(S10)}$$

Let us focus on the first two terms under the integrals. From Equation 17 of the main text, we know that $\mathrm{P}(w \mid \mu_s, s)$ is normally distributed with mean $Q(s;\, \mu_s,\, \sigma_s)$ and variance σ_c^2. In addition, from Equation 1 of the main text, we know that $\mathrm{P}(s \mid \mu_s)$ is normally distributed with mean μ_s and variance σ_s^2. The product of these two probability density functions is the bivariate normal distribution $f(\mu_s, s \mid w)$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. After a bit of algebra, we get

$$\begin{cases} \boldsymbol{\mu} = \left(w\, \frac{\sigma_s}{\beta} + \theta_s + \theta_c\right)[1,\ 1] \\ \boldsymbol{\Sigma} = \left(\frac{\sigma_s}{\beta}\right)^2 \begin{bmatrix} \sigma_c^2 + (1-\alpha)^2\beta^2 & \sigma_c^2 - (1-\alpha)\alpha\beta^2 \\ \sigma_c^2 - (1-\alpha)\alpha\beta^2 & \sigma_c^2 + \alpha^2\beta^2 \end{bmatrix} \end{cases} . \quad \text{(S11)}$$

An example of this bivariate normal distribution $f(\mu_s, s \mid w)$ is shown in Figure A2. The perceptual mode *M* is 'R' to the right of the sensory criterion (vertical dashed line) and 'L' to its left. In addition, perceptual decisions are 'R' above the sensory criterion (horizontal dashed line) and 'L' below. Therefore, self-consistent perceptual decisions are located in the upper-right and lower-left of this plot.

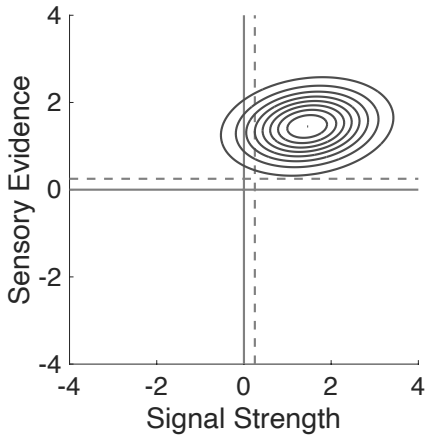

Figure A2. Joint distribution of signal strength and sensory evidence for a given confidence evidence. In this plot, confidence evidence is $w = 1.2$ (see green arrow in Figure 7) and other parameter values are those listed in Table 2.

When the stimulus strengths have a uniform probability of occurrence, Equation S10 becomes

$$\mathrm{P}(\text{self-consistent} \mid w, D) = \begin{cases} \dfrac{\sum_{\mu_S > \theta_S}\left(\int_{\theta_s}^{+\infty} f(\mu_S, s \mid w)\, ds\right)}{\sum_{\mu_S}\left(\int_{\theta_s}^{+\infty} f(\mu_S, s \mid w)\, ds\right)} & \text{if } D = \text{'R'} \\ \dfrac{\sum_{\mu_S \leq \theta_S}\left(\int_{-\infty}^{\theta_s} f(\mu_S, s \mid w)\, ds\right)}{\sum_{\mu_S}\left(\int_{-\infty}^{\theta_s} f(\mu_S, s \mid w)\, ds\right)} & \text{otherwise} \end{cases} . \quad \text{(S12)}$$

Figure A3 shows simulations of the probability of being self-consistent for different values of confidence evidence separately for each perceptual decision. As expected, this probability grows with the magnitude of the confidence evidence. Maybe more curious is the fact that the probability of being self-consistent is not exactly 0.5 when confidence evidence is zero. From Equation S11, we see that for this to happen, we need to have $\sigma_c^2 = (1-\alpha)\alpha\beta^2$ (and so, in particular, this happens for the ideal and super-ideal confidence observers). The functions for each perceptual decision are symmetric about zero confidence evidence (although small deviations can occur if the presented stimulus strengths are not symmetric about the sensory criterion), which motivates the definition of the signed confidence evidence (Equation 19 of the main text).

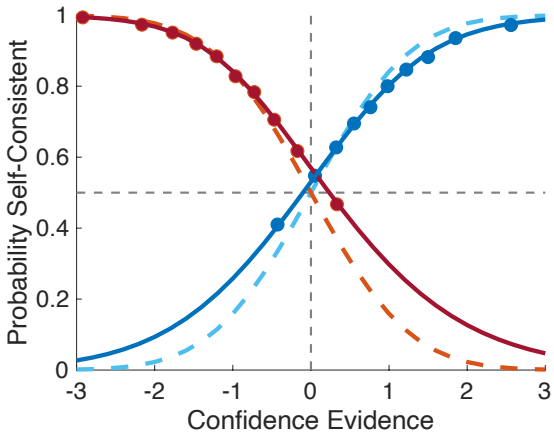

Figure A3. Relationship between self-consistency and confidence evidence. Dots represent the same 100,000 simulated trials as in Figure A1. The perceptual outcome is either 'R' (blue dots) or 'L' (red dots), but now binned in 10 bins of confidence evidence of equal size. The dark blue and red continuous curves show the probability of being self-consistent (Equation S12) obtained in the conditions of the simulations (7 stimulus strengths). The light blue and red dashed curves show the confidence probability (Equation S13) also obtained in the conditions of the simulations, but this time considering only the signed confidence evidence. In this plot, parameter values are those listed in Table 2.

In the main text, we defined the confidence probability as (see Equation 20 of the main text)

$$\mathrm{P}(\mathrm{confident} \mid w, D) = \Phi(\mathrm{w}') \quad . \tag{S13}$$

This confidence probability was offered as a proxy for the probability of being self-consistent for the current value of confidence evidence and the current perceptual decision. Confidence probability is shown in Figure A3 as light colour dashed curves. In this particular example, there is a good agreement between the confidence probability and the probability of being self-consistent, because confidence boost was close to zero ($\alpha = 0.2$) and confidence noise was moderate ($\sigma_c = 0.5$). In fact, for the ideal confidence observer (in tasks where stimuli can take any stimulus strength with uniform probability), the probability of being self-consistent equals the confidence probability. When the observer behaves a bit more like the super-ideal confidence observer, or when confidence noise is large, significant deviations can be observed between the probability of being self-consistent and the confidence probability, although the two keep a monotonic relationship.

Appendix B: Effects of interval bias

In Figure 9 of the main text, we defined the regions of the space of confidence evidence where the confidence choice was in favor of interval 1 or 2. When there is a bias in favor of the first or second interval, these regions are growing or shrinking (see Figure B1).

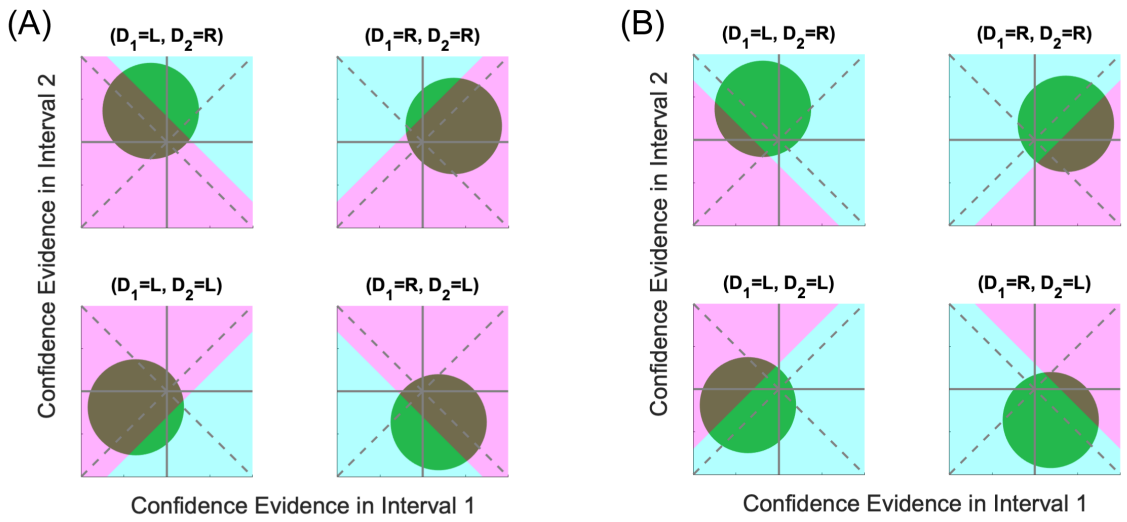

Figure B1. Confidence decision rule including interval bias. Conventions are identical to those used in Figure 9. (A) Bias in favor of interval 1. (B) Bias in favor of interval 2.

Looking at the space of sensory evidence, an interval bias also affects the confidence choice map (Figure B2).

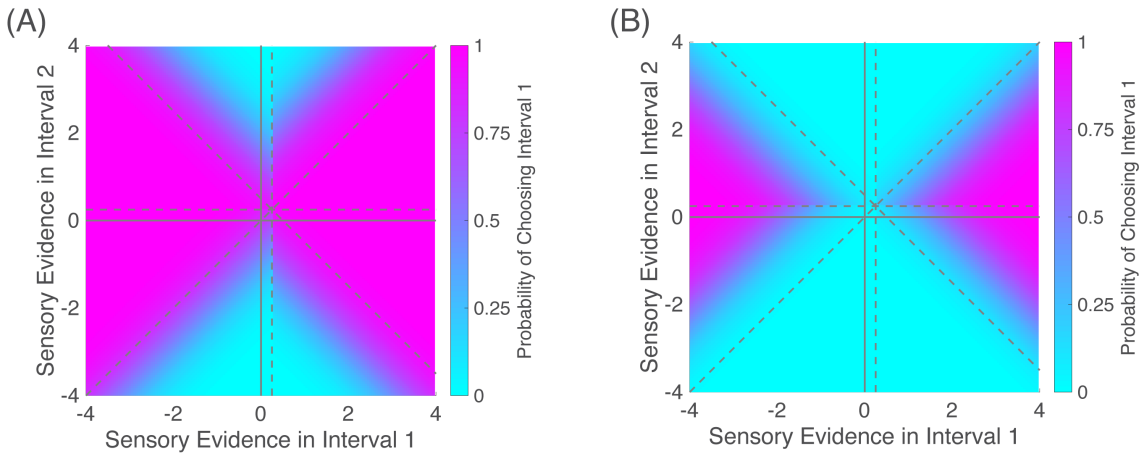

Figure B2. Confidence choice map. The probability of choosing interval 1 as more confident is plotted for each pair of sensory evidences in intervals 1 and 2. Parameters used to generate the figures are those listed in Table 2 except for a bias

in favor of the first interval ($\gamma = +1$; panel A) or a bias in favor of the second interval ($\gamma = -1$; panel B). Figure conventions are identical to those used for Figure 11B.

Appendix C: Effects of confidence boost and confidence noise

We will look at the effect of the confidence boost and the confidence noise parameters on two critical plots, the joint distribution of sensory and confidence evidences $H(s,w)$ (see Figure 7) and the confidence choice map (see Figure 11B).

We first look at the effect of the confidence boost (Figure C1). As the confidence boost increases, the confidence evidence becomes more independent of the sensory evidence (see Equation 13 of the main text). This results in a reduced variance of the confidence evidence (see Equation 11 of the main text), and a level of confidence evidence that is more likely to be near the mean of the distribution (see Equation 10 of the main text). As a consequence, as the confidence boost increases, one interval becomes more and more likely that its sensory evidence will determine which interval will be chosen as the more confident, irrespective of the sensory evidence in the other interval. This winning interval is the one for which the mean of the confidence evidence is larger in absolute value. In the example of Figure C1, the means of confidence evidence in intervals 1 and 2 are respectively 1.25 $(1.5 - 0.25)$ and –0.75 $(-0.5 - 0.25)$. Therefore, the interval whose sensory evidence determines the confident interval is here interval 1 (this creates the vertical segregation in Figure C1H).

Next, we look at the effect of the confidence noise (Figure C2). As the confidence noise increases, the confidence evidence covers an increasing range of values. In addition, as the confidence noise increases, the two intervals have a more similar probability of being chosen as the more confident one.

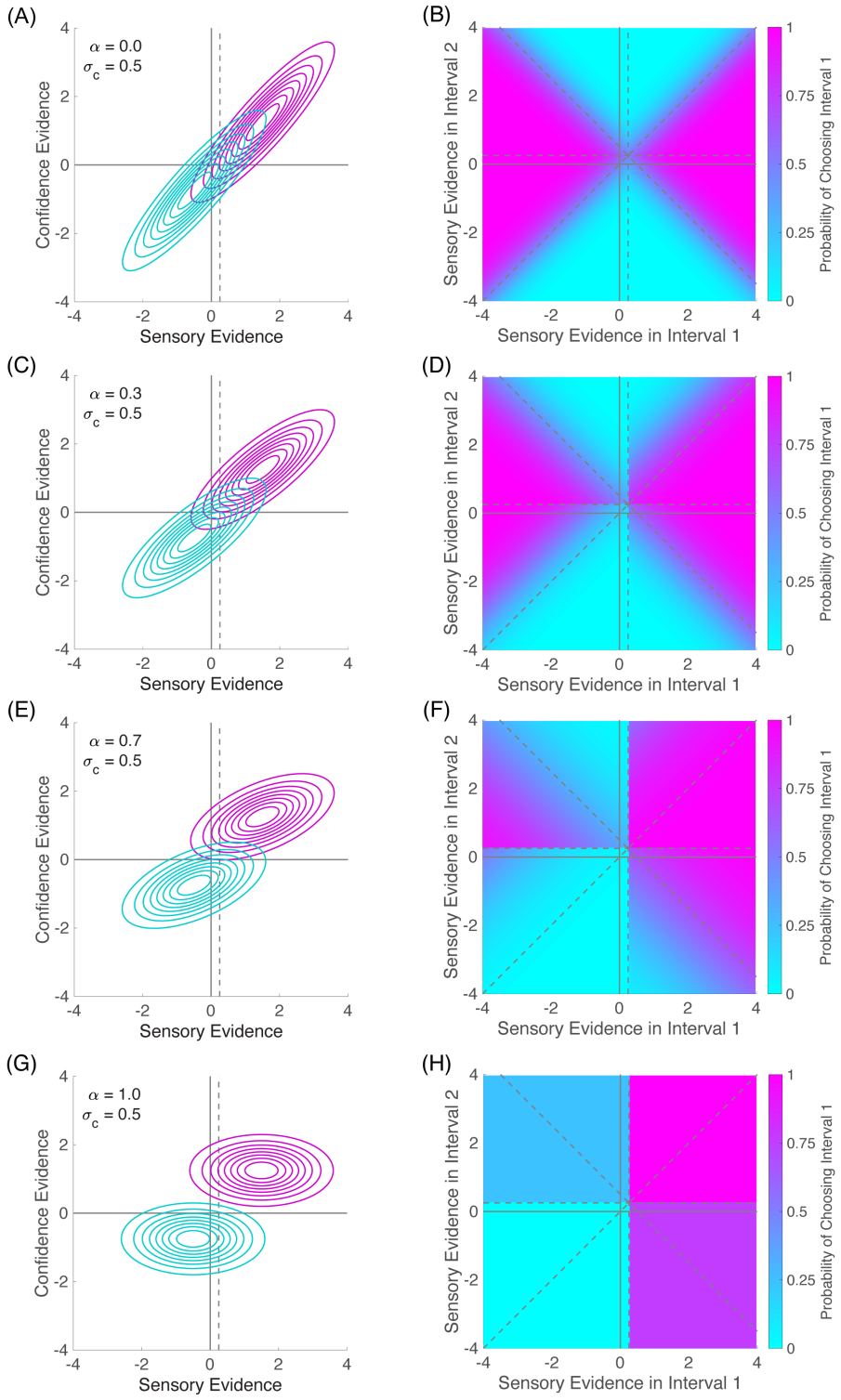

Figure C1. Effect of confidence boost on joint distributions. The left column (A, C, E, and G) shows the joint distribution of sensory and confidence evidences for different values of the confidence boost. The joint distributions are shown for both interval 1 (purple) and interval 2 (cyan). The right column (B, D, F, and H) shows the confidence choice map for the corresponding values of the confidence boost in the left column. (A, B) $\alpha = 0$. (C, D) $\alpha = 0.33$. (E, F) $\alpha = 0.67$. (G, H) $\alpha = 1$. All other parameters are those listed in Table 2.

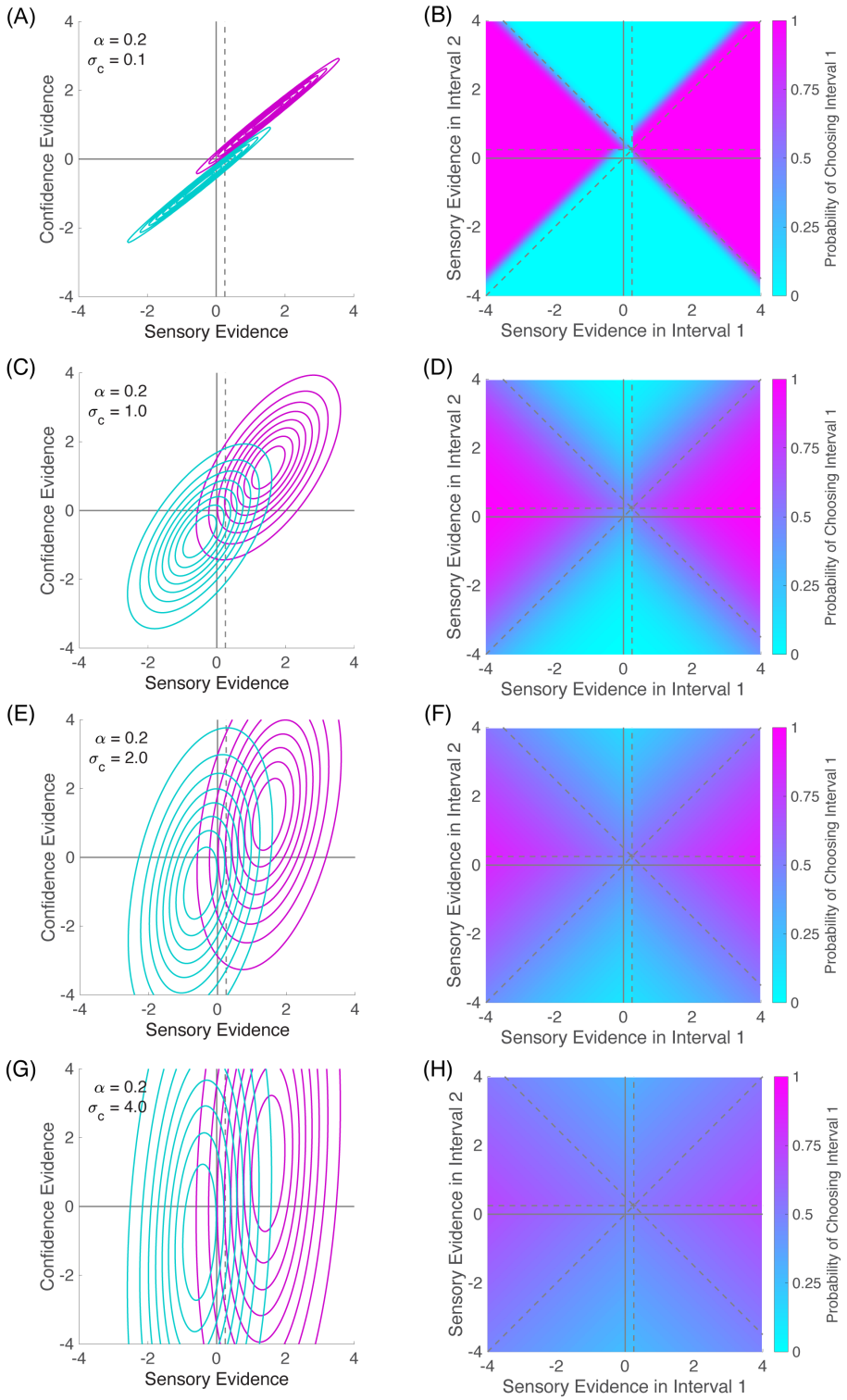

Figure C2. Effect of confidence noise on joint distributions. The left column (A, C, E, and G) shows the joint distribution of sensory and confidence evidences for different values of the confidence noise. The joint distributions are shown for both interval 1 (purple) and interval 2 (cyan). The right column (B, D, F, and H) shows the confidence choice map for the corresponding values of the confidence noise in the left column. (A, B) $\sigma_c = 0.1$. (C, D) $\sigma_c = 1.0$. (E, F) $\sigma_c = 2.0$. (G, H) $\sigma_c = 4.0$. All other parameters are those listed in Table 2.

Appendix D: Parameter recovery

We present a set of simulations to show that all the parameters of the model can be recovered, and also that each parameter does not introduce biases on the other parameters. The first two parameters we consider are the sensory noise (Figure D1) and the sensory criterion (Figure D2).

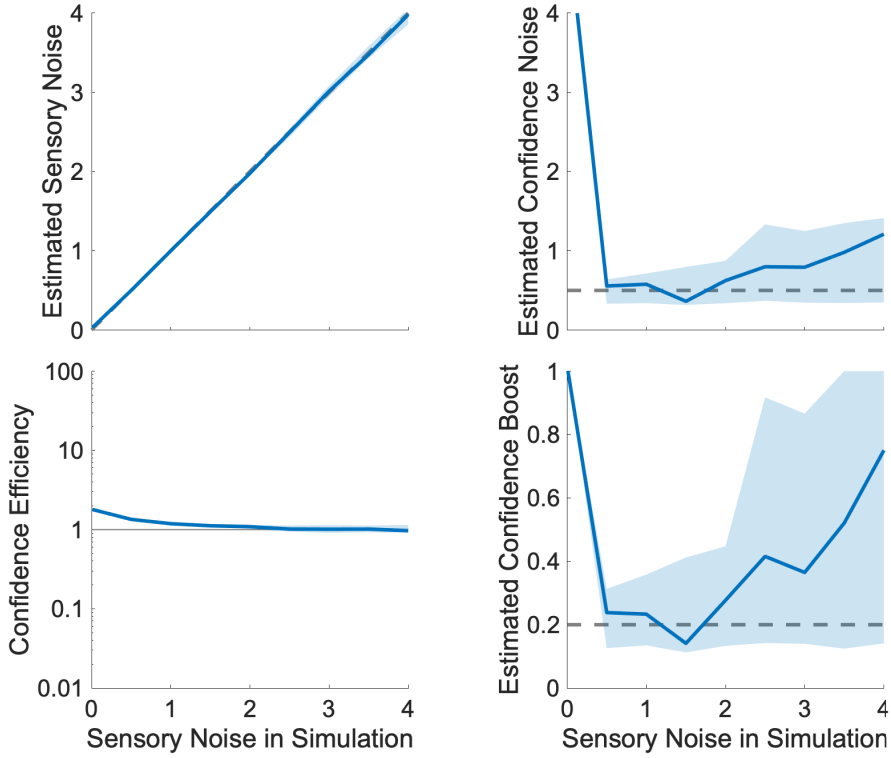

Figure D1. Parameter recovery for different sensory noise. The plots show estimated parameters for different values of the parameter σ_s that corresponds to the sensory noise. In each plot, the thick lines are median estimated values across 20 repeated simulations, and the shaded areas cover the 25th to the 75th interquartile range. The estimated parameters are the sensory noise (top-left panel), confidence noise (top-right), and confidence boost (bottom-right). Estimated confidence efficiency is stable for different sensory noise (bottom-left panel). The estimated sensory noise is well recovered from the original psychometric function. Confidence noise and confidence boost are well estimated only when the range of stimulus strengths is appropriate for the sensory noise (i.e. proportion self-consistent varies from chance to almost ceiling performance). When sensory noise is zero, there is no variability in the perceptual response, and thus there is no room to estimate the confidence boost and confidence noise.

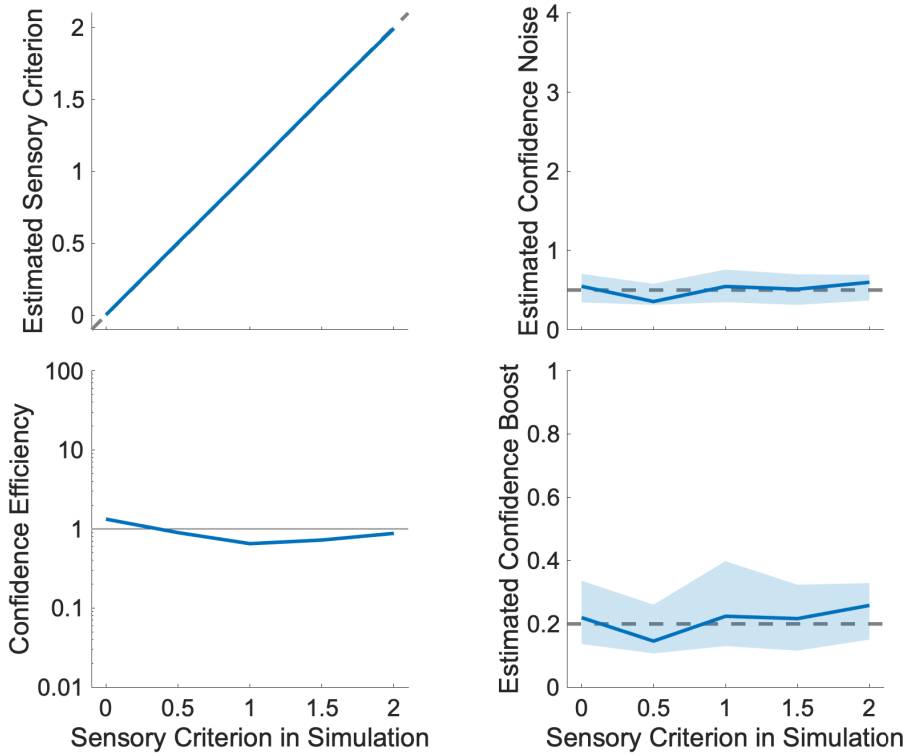

Figure D2. Parameter recovery for different sensory criteria. The plots show estimated parameters for different values of the parameter θ_s that corresponds to the sensory criterion. The thick lines are median estimated values across 20 repeated simulations, and the shaded areas cover the 25th to the 75th interquartile range. The estimated parameters are the sensory criterion (top-left panel), confidence noise (top-right), and confidence boost (bottom-right). Confidence efficiency is stable across sensory criteria (bottom-left panel). The estimated sensory criterion is well recovered from the original psychometric function.

The next parameter is a confidence criterion that is potentially distinct from the sensory criterion (Figure D3).

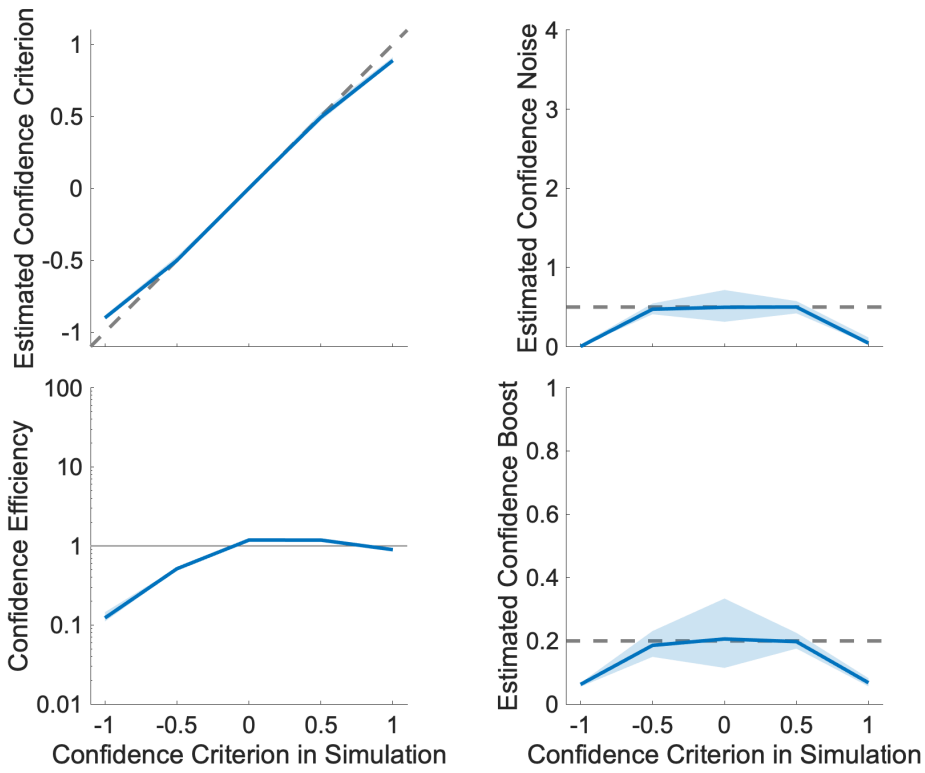

Figure D3. Parameter recovery for different confidence criteria. The plots show estimated parameters for different values of the parameter θ_c that corresponds to the confidence criterion. The thick lines are median estimated values across 20 repeated simulations, and the shaded areas cover the 25th to the 75th interquartile range. The estimated parameters are the sensory criterion (top-left panel), confidence noise (top-right), and confidence boost (bottom-right). These parameters are well recovered only within the limits of the range of visited sensory stimuli. Confidence efficiency (bottom-left panel) decreases when the difference between confidence and sensory criteria increases.

One important parameter is a potential bias to choose one interval rather than the other (Figure D4).

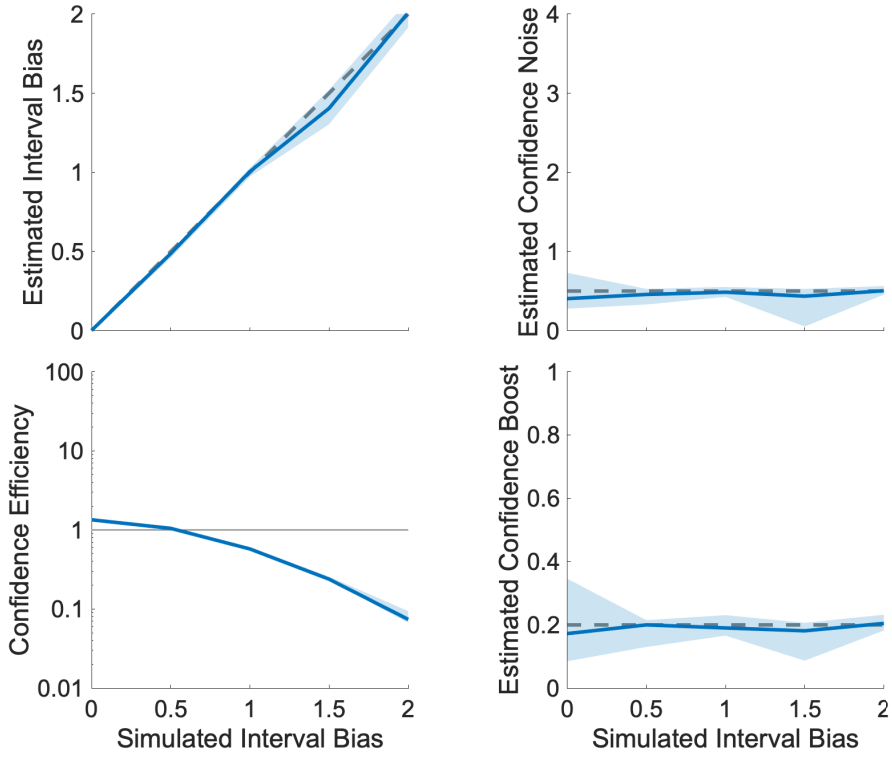

Figure D4. Parameter recovery for different interval biases. The plots show estimated parameters for different values of the parameter γ that reflects biases in favour of the first interval. The thick lines are median estimated values across 20 repeated simulations, and the shaded areas cover the 25th to the 75th interquartile range. The estimated parameters are the interval bias (top-left panel), confidence noise (top-right), and confidence boost (bottom-right). Estimated confidence efficiency decreases when the interval bias increases (bottom-left panel). The estimated interval bias is well recovered.

The next set of simulations shows how well we can recover the confidence noise and confidence boost parameters when the number of confidence pairs varies (Figure D5). As expected, confidence parameters are better estimated as the number of trials increases. Below 1,000 confidence pairs, the estimated slope gain is relatively uncertain (less precise) whereas confidence efficiency is underestimated (less accurate).

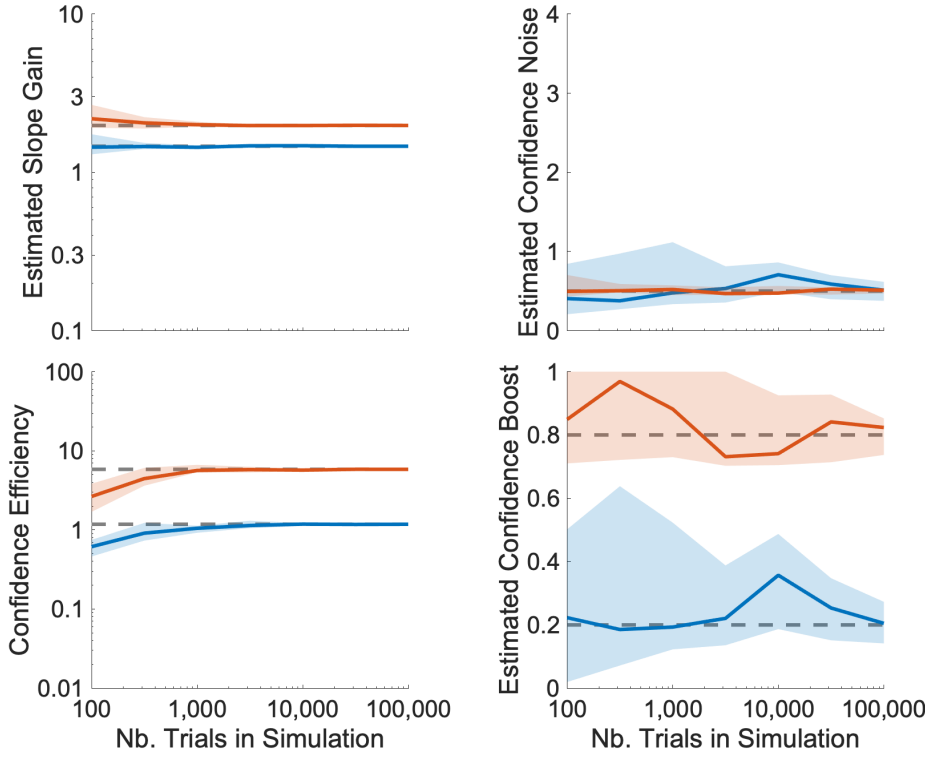

Figure D5. Parameter recovery for different numbers of confidence pairs. The plots show estimated parameters for two different values of confidence boost, $\alpha = 0.2$ in blue and $\alpha = 0.8$ in orange, all other parameters are those listed in Table 2. The gain in the slope of the chosen psychometric function and the confidence efficiency are shown in the left column, while the estimated confidence noise and confidence boost are shown in the right column. The thick lines are median estimated values across 20 repeated simulations, and the shaded areas cover the 25th to the 75th interquartile range. Dashed lines in the left column indicate the asymptotic values for large number of simulated trials, whereas in the right column, they indicate the values of the parameters used in the simulations.

The final set of simulations shows how well we can recover the different parameters of the model when the range of the stimulus strengths varies (Figure D6). We note that confidence efficiency and slope gain are very stable over different ranges of stimulus strengths, including when there is a single stimulus strength (i.e. when the range is zero). In contrast, the confidence noise and confidence boost are well estimated only when the range of stimulus strengths is relatively wide. In the extreme scenario where there is a single stimulus strength, the confidence boost parameter is indeterminate. This is the exact case of confidence parameters indeterminacy described in the text.

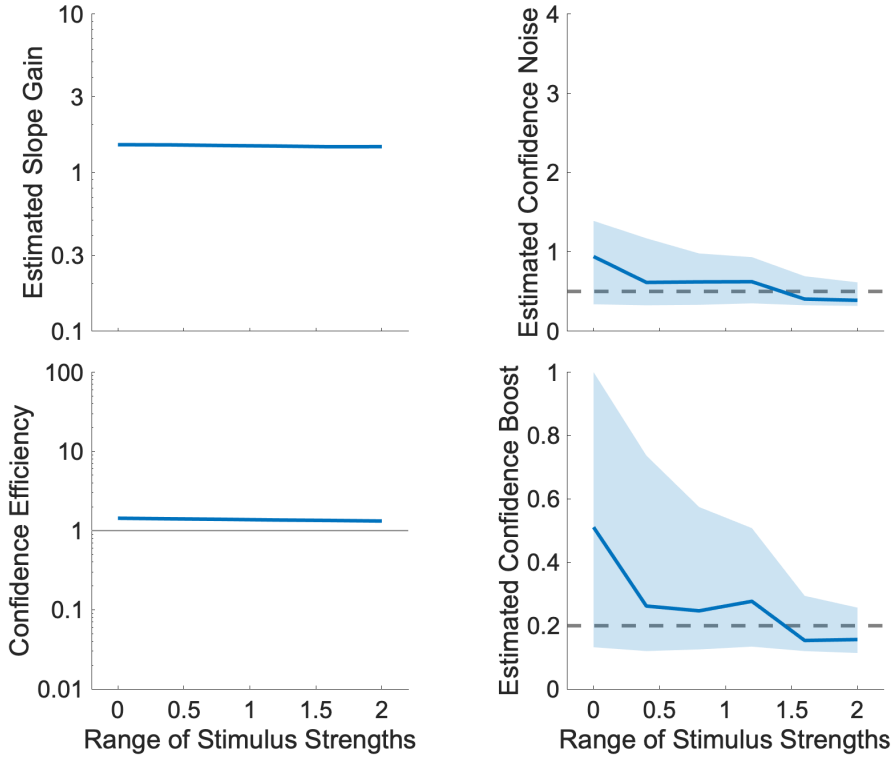

Figure D6. Parameter recovery for different ranges of stimulus strengths. Ranges of stimulus strengths are ranges of 4 values over the nominal values of -0.75 or 0.75 (e.g. when the range is 0.4, the simulated stimulus strengths are [-0.9, -0.8, -0.7, -0.6, 0.6, 0.7, 0.8, 0.9]). All other parameters are those listed in Table 2 (the range in that table is 2.0), except that the sensory criterion was set to 0 (otherwise we would not reach a single actual stimulus strength in the extreme scenario where the range was 0). The gain in the slope of the chosen psychometric function and the confidence efficiency are shown in the left column, while the estimated confidence noise and confidence boost are shown in the right column. The thick lines are median estimated values across 50 repeated simulations, and the shaded areas cover the 25th to the

75th interquartile range. Dashed lines in the right column indicate the values of the parameters used in the simulations.

Appendix E: Recovery of parameters when there are two tasks

When two distinct tasks are present across the two intervals of a confidence pair, we need to estimate both confidence boosts for the two tasks and both confidence noises. We present here results from simulations where either task could be presented in either interval. If the confidence trials only consist in comparisons across tasks, we found in simulations that the confidence boost could be accurately estimated, but only one of the two confidence noises could be estimated (the second one is anti-correlated with the first one). When either task could be presented in either interval, both confidence noises and both confidence boosts can be accurately recovered (Figures E1 and E2).

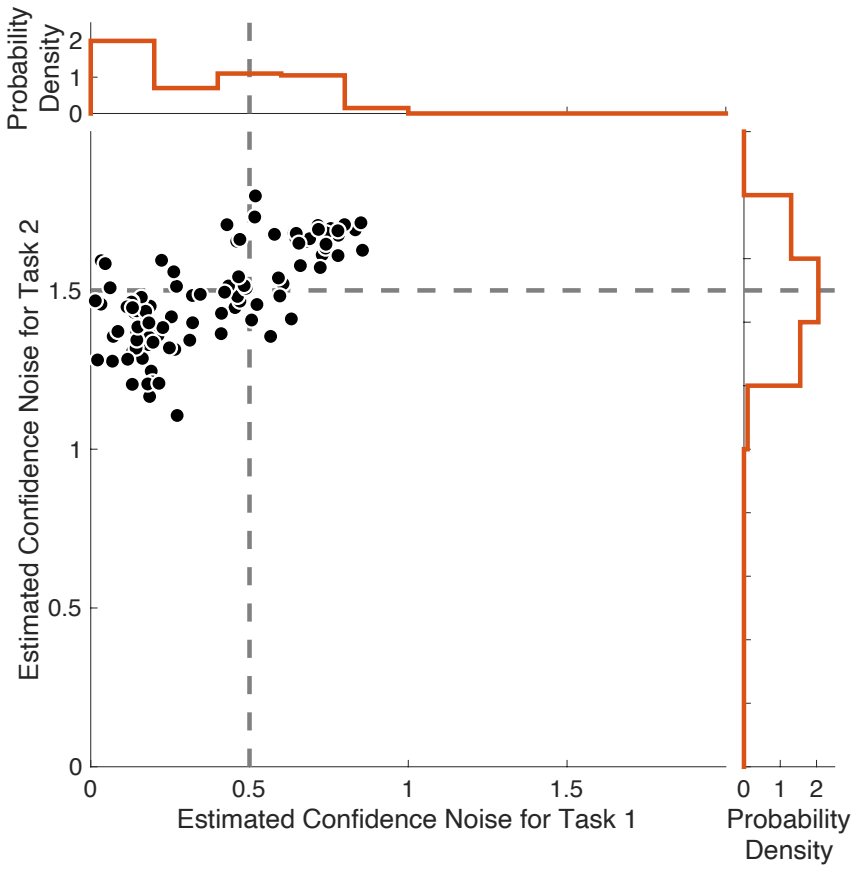

Figure E1. Parameter recovery for confidence noise when there are two tasks. The simulated model is shown in dashed lines, and the parameters were $\sigma_c = 0.5$ and $\alpha = 0.2$ for task 1, and $\sigma_c = 1.5$ and $\alpha = 0.8$ for task 2. The central plot shows 100 simulations of the model, and the upper and right panels show marginal distributions for the estimated confidence noise for task 1 and task 2, respectively. The other parameters are those listed in Table 2.

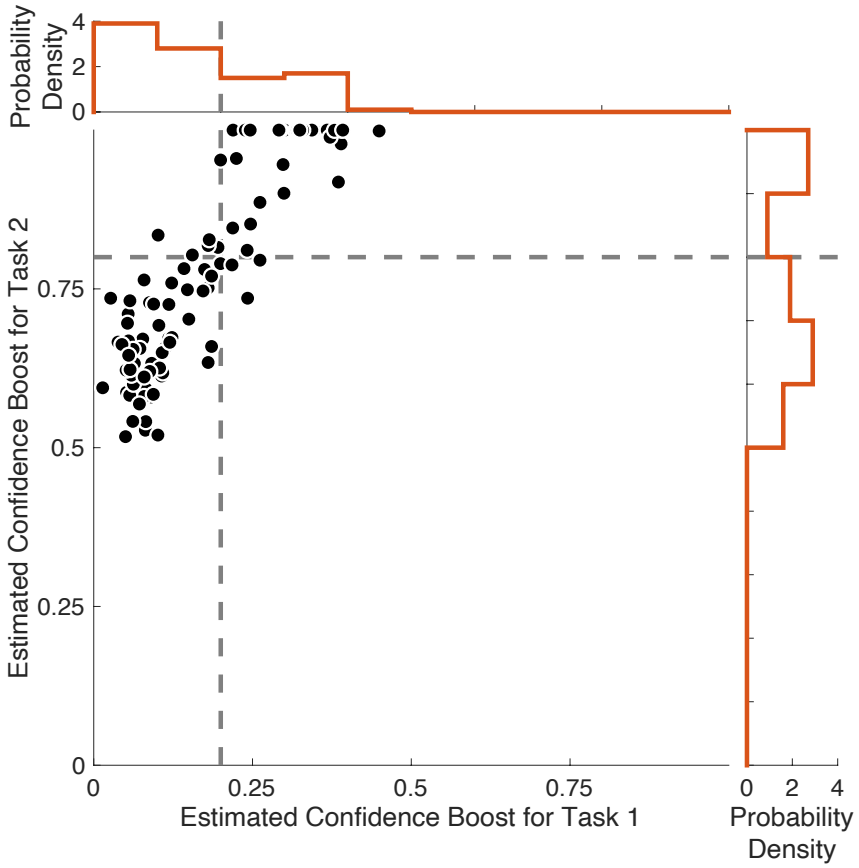

Figure E2. Parameter recovery for confidence boost when there are two tasks. The simulated model is shown in dashed lines, and the parameters were $\sigma_c = 0.5$ and $\alpha = 0.2$ for task 1, and $\sigma_c = 1.5$ and $\alpha = 0.8$ for task 2. The central plot shows 100 simulations of the model, and the upper and right panels show marginal distributions for the estimated confidence boost for task 1 and task 2, respectively. The other parameters are those listed in Table 2.

When the experiment involves two distinct tasks, we can estimate the confidence bias of one task relative to the other (Figure E3).

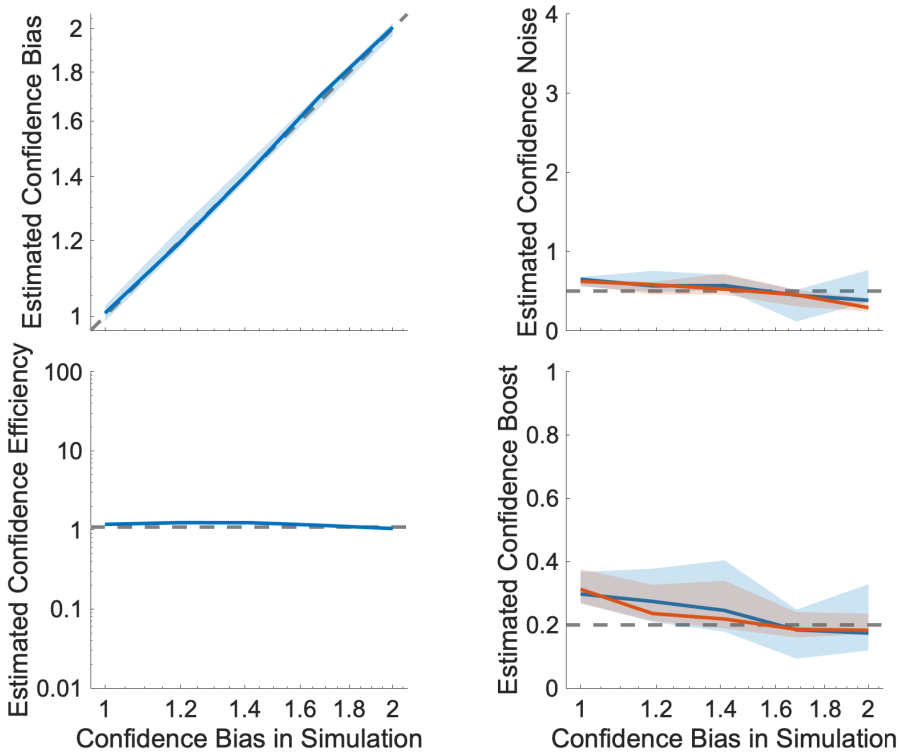

Figure E3. Parameter recovery for a range of confidence bias. The top-left plot shows the estimated confidence bias for a range of simulated parameters, from no confidence bias ($\beta = 1$) to an overconfidence for task 2 relative to task 1 ($\beta > 1$). The estimated confidence efficiency is stable across variations of confidence bias. Confidence noise and confidence boost parameters (right column) are also relatively well estimated for both the first (blue) and second (orange) tasks. Thick lines are median estimated values across 20 repeated simulations, and shaded areas cover the 25th to the 75th interquartile range.

Appendix F: Detailed generative model

In this section, we present the derivations of the full model that includes both parallel processing of confidence and accumulated sensory evidence after the perceptual decision has been taken. In the main manuscript, these two sources of information were conflated into what we called secondary confidence evidence. We justify this choice here arguing that it is difficult to disentangle these two sources.

Following some conventions similar to the ones in the main text, we define three sources of confidence evidence. First, the ideal confidence observer is still defined as the observer that uses exactly the same evidence as the one that was used to reach the perceptual decision. Therefore,

$$w_{\text{ideal}} = (s - \theta_s)/\sigma_s = (\mu_s + \epsilon_s - \theta_s)/\sigma_s \quad . \tag{S14}$$

The second source of confidence evidence is coming from a parallel intake of new sensory information from the stimulus. It is subject to its own sensory criterion θ_s' and sensory noise σ_s', and to its own sample of sensory noise ϵ_s' drawn from a zero-mean normal distribution with standard deviation σ_s'. Therefore, the parallel confidence evidence is defined as

$$w_{\text{parallel}} = (\mu_s + \epsilon_s' - \theta_s')/\sigma_s' \quad . \tag{S15}$$

The third source of confidence evidence is coming from serial information accumulated after the perceptual decision. It is therefore relying on the same sensory apparatus (and therefore submitted to the same sensory sensitivity and criterion), but with a different sample of sensory noise ϵ_s'' drawn from a zero-mean normal distribution with standard deviation σ_s. Therefore, this serial (or post-decisional) confidence evidence is defined as

$$w_{\text{serial}} = (\mu_s + \epsilon_s'' - \theta_s)/\sigma_s \quad . \tag{S16}$$

To these sources of confidence evidence that are directly linked to the current stimulus, we can add a fourth source that includes everything else. This source might reflect variations of confidence due to for example fluctuations of attention, as well as some information contained in response times above and beyond the information coming from the ideal evidence. Because this source is unrelated to the current stimulus, we can arbitrarily set its evidence to a fixed value of 1

$$w_{\text{other}} = 1 \quad . \tag{S17}$$

Altogether, the confidence evidence of one observer for one particular perceptual decision will be a weighted sum of all these sources of evidence. To this weighted combination, we also consider some additive confidence noise ϵ_c' drawn from a zero-mean normal distribution with standard deviation σ_c'

$$w = \alpha_1\, w_{\text{ideal}} \;+\; \alpha_2\, w_{\text{parallel}} \;+\; \alpha_3\, w_{\text{serial}} \;+ \alpha_0\, w_{\text{other}} + \;\epsilon_c' \;\; . \tag{S18}$$

In the main text, we have proposed that the confidence evidence takes the following general form (see again Equation 9)

$$w = \; (\mu_s + \; (1-\alpha)\,.\,\epsilon_{\text{s}} \;-\; \theta_{\text{s}} - \; \theta_{\text{c}})\,.\,\beta/\sigma_{\text{s}} \;+\; \epsilon_{\text{c}} \;\; . \tag{S19}$$

We can make the link between Equations S18 and S19 by developing each term Equation S18. From the definitions above, we obtain

$$w = \alpha_1\,(\mu_s + \epsilon_{\text{s}} - \theta_s)/\sigma_s \;+\; \alpha_2\,(\mu_s + \epsilon_s' - \theta_s')/\sigma_s' \;+\; \alpha_3\,(\mu_s + \epsilon_s'' - \theta_s)/\sigma_s \;+ \alpha_0 + \;\epsilon_c' \;\; . \tag{S20}$$

Grouping together the terms in μ_s and in ϵ_{s}, we obtain

$$w = \mu_s \left(\frac{\alpha_1}{\sigma_s} + \frac{\alpha_2}{\sigma_s'} + \frac{\alpha_3}{\sigma_s}\right) + \epsilon_{\text{s}} \left(\frac{\alpha_1}{\sigma_s}\right) + \epsilon_s' \left(\frac{\alpha_2}{\sigma_s'}\right) + \epsilon_s'' \left(\frac{\alpha_3}{\sigma_s}\right) + \left(\alpha_0 - \frac{\alpha_1\theta_s}{\sigma_s} - \frac{\alpha_2\theta_s'}{\sigma_s'} - \frac{\alpha_3\theta_s}{\sigma_s}\right) + \;\epsilon_c' \;. \tag{S21}$$

In our model, because confidence was compared across two consecutive perceptual trials, any scaled version of this confidence evidence will be appropriate. We choose as scaling factor

$$A = \alpha_1 + \;\alpha_2\,\frac{\sigma_s}{\sigma_s'} + \alpha_3 \;\; , \tag{S22}$$

and define the new weight α, in the interval $[0, 1]$, as

$$\alpha = 1 - \frac{\alpha_1}{A} \;=\; \frac{\alpha_2(\sigma_s/\sigma_s') + \alpha_3}{\alpha_1 + \alpha_2(\sigma_s/\sigma_s') + \alpha_3} \;\; , \tag{S23}$$

such that

$$w = \frac{\mu_s}{\sigma_s} + \frac{(1-\alpha)\,\epsilon_{\text{s}}}{\sigma_s} \;+\; \left(\alpha_0 - \frac{\alpha_1\theta_s}{\sigma_s} - \frac{\alpha_2\theta_s'}{\sigma_s'} - \frac{\alpha_3\theta_s}{\sigma_s}\right)\frac{1}{A} \;+\; \frac{\alpha_2\,\epsilon_s'}{\sigma_s'} + \;\frac{\alpha_3\,\epsilon_s''}{\sigma_s} + \frac{\epsilon_c'}{A} \;\; . \tag{S24}$$

The last three terms are random variables that can be grouped into a single zero-mean normal distribution with standard deviation σ_{c}. The implication is that the samples of sensory noise specific to the parallel and serial streams ϵ_s' and ϵ_s'' are absorbed in the confidence noise.

The constant term in the middle of Equation S24 can be equated to the criteria in Equation S19

$$\theta_{\text{s}} + \;\theta_{\text{c}} = -\left(\alpha_0 - \frac{\alpha_1\theta_s}{\sigma_s} - \frac{\alpha_2\theta_s'}{\sigma_s'} - \frac{\alpha_3\theta_s}{\sigma_s}\right)\frac{\sigma_s}{A} \;\; , \tag{S25}$$

so that

$$\theta_{\text{c}} = \frac{[\alpha_2(\theta_s' - \theta_s) - \alpha_0\sigma_s']\,\sigma_s}{A\,\sigma_s'} \;\; . \tag{S26}$$

Finally, if we wanted to also consider a confidence bias β in misestimating the sensory noise of the ideal confidence observer, we would need to replace σ_s by σ_s/β in the above expressions. In summary, the interpretation of the parameters of our model is the following.

Let us start with the confidence boost α. If α is zero, then both α_2 and α_3 are also zero, so that the observer does not rely on secondary confidence evidence, and, if other sources of confidence are neglected ($\alpha_0 = 0$), the observer relies only on primary confidence evidence. On the contrary, if α is one, then α_1 is zero, indicating that the observer relies on secondary confidence evidence and not on primary evidence.

The confidence noise σ_c is actually a combination of additive confidence noise and sensory noises from the secondary stream of confidence computation, including both parallel and serial post-decisional components.

The interpretation of the confidence criterion θ_c is a bit more complex. In practice, we expect θ_s' to be close to θ_s, and other sources of confidence to be small ($\alpha_0 = 0$). Therefore, we do not expect this parameter to be playing a major role, and we can avoid including it in the fitting procedure in a first pass.

Appendix G: Application of the model to confidence ratings

In this section, we show that our generative model can also be applied to data collected with a confidence rating paradigm. We consider the simple scenario where stimuli have a single strength and confidence is judged on a binary scale. This confidence judgment involves a single confidence rating criterion, so that any confidence evidence above this criterion is judged to be high confidence, and below the criterion to be low confidence.

We first look at the joint distribution of sensory and confidence evidence (Figure G1A). Because the stimulus strength ($\mu_s = 1.5$) was above the sensory criterion ($\theta_s = 0.25$), self-consistent perceptual decisions are to the right of the sensory criterion (shown in blue), and self-inconsistent decisions are to its left (shown in red). The marginal distributions of confidence evidence for self-consistent and self-inconsistent decisions are shown in the panel on the right of Figure G1A.

Confidence judgments in a rating task consist in comparing the signed confidence evidence w' to a confidence rating criterion. The signed confidence evidence is simply the confidence evidence for self-consistent perceptual decisions, and the opposite of confidence evidence for self-inconsistent decisions. Therefore, high confidence judgements correspond to confidence evidence above the confidence rating criterion for self-consistent perceptual decisions, and below the opposite of the rating criterion for self-inconsistent decisions. These are shown as shaded areas in the marginal distributions.

Figure G1B shows the Type 2 Receiving Operating Characteristic (ROC) curve. The Type 2 hit rate is the probability of making a high confidence judgment when the perceptual decision was self-consistent, and therefore is the blue shaded area in Figure G1A. Similarly, the Type false alarm is the probability of making a high confidence judgment when the perceptual decision was self-inconsistent, and therefore is the red shaded area in Figure G1A. Because there is a single stimulus strength and a single rating criterion, there is a single point in the Type 2 ROC curve shown as the open circle in Figure G1B.

The point in the Type 2 ROC curve with coordinates (Type 2 false alarm rate, Type 2 hit rate) depends on the simulated values for the confidence boost $\alpha = 0.2$, the confidence noise $\sigma_c = 0.5$, and the confidence rating criterion $k = 0.8$. Importantly, the same point can be obtained for other combinations of confidence boost and confidence noise. For instance, setting the confidence boost to 1, the same point can be obtained with a confidence noise of 1.9 (what we called the equivalent confidence noise in the main text), and a confidence rating criterion of 0.29 (other rating criteria produce the green dashed line in Figure G1B). This property of obtaining the same confidence judgments for different combinations of confidence boost and confidence noise is what we called confidence parameters indeterminacy in the main text.

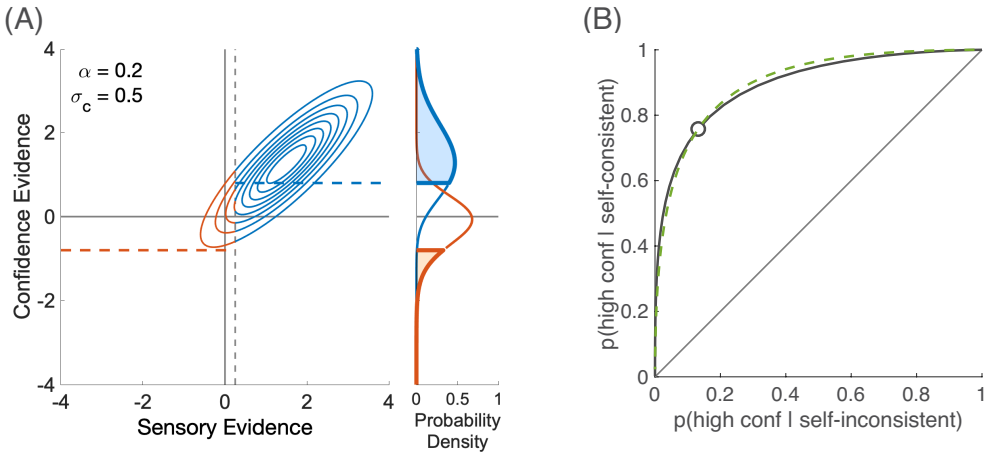

Figure G1. Application of the model to confidence rating. (A) Joint distribution of sensory and confidence evidence. Blue and red colors correspond to self-consistent and self-inconsistent perceptual decisions, respectively. The sensory criterion is shown as the grey vertical dashed line. The confidence rating criterion was set to $k = 0.8$ and is shown as the colored dashed horizontal lines. All other parameters are those listed in Table 2. (B) Type 2 ROC curve. The confidence rating criterion in panel (A) correspond to the circle in this plot, and varying this rating criterion generates the grey solid line. The green dashed line corresponds to another pair of confidence boost and confidence noise that also goes through the simulated point on the Type 2 ROC curve for one particular value of confidence rating criterion (see text for details).