Supplementary Materials

Table of contents

1. Experiment 1. A proxy for local confidence based on an ideal confidence model

To investigate how global confidence would depend on local confidence for setsizes larger than 1 in Experiment 1, we relied on a proxy for local confidence. Below we describe in detail the steps we followed to build this proxy which is based on a SDT model of ideal confidence.

Perceptual sensitivity used in the SDT model of ideal confidence

Initially, perceptual sensitivity was evaluated using a typical psychophysical function relating the proportion of clockwise (*CW*) responses to the difference in degrees between the two visual orientations (*S*), using a cumulative Gaussian. However, to take into account the effects of position in the set (*Pos* between 1 and 8), and the effect of time in the experiment (*T* between 1 and 40*)*, we then included these variables in our psychophysical fit. Because time had a U-shaped relation with performance, we also added T^2 to our model. More precisely, we estimated the following equation:

$$P(CW) = \phi(\alpha_0 + S \times (\alpha_1 + \alpha_2 \times Position + \alpha_3 \times T + \alpha_4 \times T^2)$$

with ϕ the cumulative distribution function of the standard normal distribution. Fitting was done by maximizing likelihood with a probit link.

This analysis allowed us to evaluate a more precise measure of the observer's internal noise, which would be used in our SDT model of ideal confidence.

$$\sigma^{sub} = \frac{1}{\alpha_1+\alpha_2\times Position+\alpha_3\times T+\alpha_4\times T^2}$$

Ideal confidence

Then, for each decision, we derived the confidence that would be expected from an agent that would have the same perceptual performance (and more specifically the same decisions on each perceptual trial) and an ideal metacognition.

To do so, we consider the following generative model. The stimulus (S) is presented, and the observer has to judge whether it corresponds to one category (e.g. clockwise, CW, for S>0) or the other (S<0). Because of internal noise, the observer receives a signal (noted s) normally distributed around the true stimulus value S, with standard deviation σ^{sub}. In addition, the ideal observer has the correct prior about the probability of each stimulus.

Suppose that an observer with internal noise answered 'clockwise' to a trial following a signal s. In a series without variance, the observer knows that with equal probability, the stimulus is either x_{25} or x_{75}. Ideal confidence is thus:

$$c^{*}|s = P(S > 0|s) = \frac{P(s|S = x_{75})}{P(s|S = x_{25}) + P(s|S = x_{75})} = \frac{h(x_{75} - s)}{h(x_{25} - s) + h(x_{75} - s)}$$

with h the pdf of a normal Gaussian with mean 0 and sd σ^{sub}.

In series with variance, the observer knows that with equal probability, stimulus is either drawn from a normal distribution centered on x_{75} or on x_{25} with standard deviation $\sigma^{stim}=(x_{75}-x_{25})/4$. Ideal confidence is thus:

$$c^{*}|s = P(S > 0|s) = P(S > 0|(s \cap S \sim N(x_{75})) \times P(S \sim N(x_{75})|s) + P(S > 0|(s \cap S \sim N(x_{25})) \times P(S \sim N(x_{25})|s)$$

Using Bayes' rule and $P(S \sim N(x_{75})) = P(S \sim N(x_{25})) = .5$, we get:

$$c^{*}|s = \frac{P(S>0|(s \cap S \sim N(x_{75})) \times P(s|S \sim N(x_{75})) + P(S>0|(s \cap S \sim N(x_{25})) \times P(s|S \sim N(x_{25}))}{P(s|S \sim N(x_{25})) + P(s|S \sim N(x_{75}))}$$

Then, using $P(s|S \sim N(z)) = h'(z - s)$, $P(S > 0|(s \cap S \sim N(x_{75})) = (1 - F(-x_{75}))$ and $P(S > 0|(s \cap S \sim N(x_{25})) = (1 - G(-x_{75}))$, we recover an explicit formula for ideal confidence:

$$c^{*}|s = \frac{(1-F(-x_{75}))\times h'(x_{75}-s)+(1-G(-x_{25}))\times h'(x_{25}-s)}{h'(x_{75}-s)+h'(x_{25}-s)}$$

With :

- F the cdf of a normal Gaussian with mean $\frac{s-x_{75}}{\sigma^{sub^2}\times A}$ and sd $= \frac{1}{\sqrt{A}}$
- G the cdf of a normal Gaussian with mean $\frac{s-x_{25}}{\sigma^{sub^2}\times A}$ and sd $= \frac{1}{\sqrt{A}}$
  - $A = \frac{\sigma^{sub^2}+\sigma^{stim^2}}{\sigma^{sub^2}\times\sigma^{stim^2}}$
- h' the pdf of a normal Gaussian with mean 0 and sd= $\sqrt{\sigma^{sub^2} + \sigma^{stim^2}}$

Expected ideal confidence

Although we did not have access to the signal received by observers, we observed their responses which told us about the sign of the signal received: a CW response indicated s>0, and a CCW response indicated s<0. We could then compute the expected ideal confidence at each trial as:

- $\int_{0}^{+Inf} P(S > 0|s)ds$ for a CW response clockwise or
- $\int_{-Inf}^{0} P(S < 0|s)$ otherwise

Local confidence proxy: mapping expected ideal confidence to confidence

Finally, for each participant, we fitted this expected ideal confidence to the confidence ratings for setsize 1, with 2 free parameters to account for a linear mapping between

expected ideal confidence and participant's reported confidence. Figure S1 shows the relationship between predicted local confidence (our proxy) and actual local confidence for sets of size 1. The R-squared across the whole dataset is 24.54%.

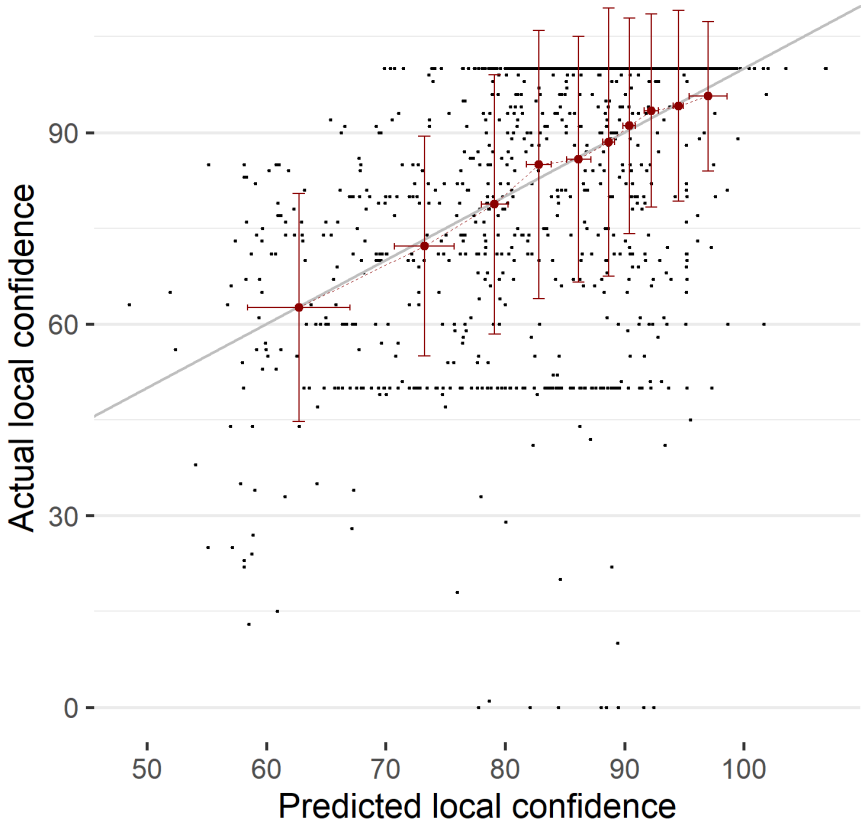

Figure S1. Actual local confidence reported in sets of size 1 against the position of the trial within the set, separately for each setsize. Each black dot represents an observation for a given participant (72 local confidence by individuals). Red error bars represent mean and s.d. over a decile of predicted local confidence.

2. Experiment 1. Effect of variability on overconfidence

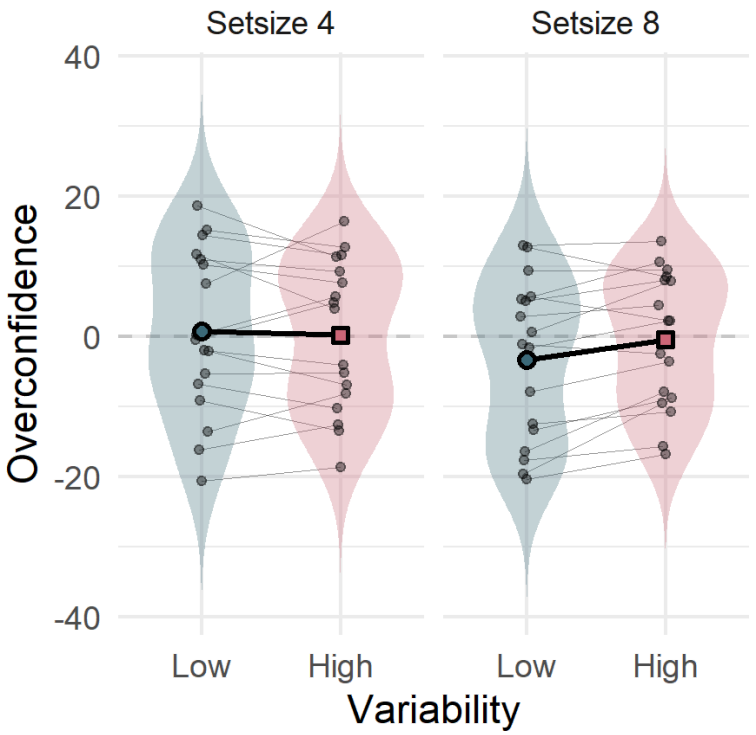

Figure S2. Distribution of overconfidence across participants in sets of size 4 and 8 in low and high variability. Large dots and thick lines represent the mean overconfidence across participants. Small dots and thin lines represent individual data.

3. Experiment 3. Recency effects in the numerical averaging task

In Experiment 3, we checked whether the recency effects on global confidence found in Experiment 1 and 3 were also present in the numerical averaging task. We thus estimated the weight of each number on participants' reported average in the numerical averaging task, and evaluated how these weights differed between positions within the sequence of numbers (between 1 and 6). An ANOVA on the weights indicated a main effect of position ($F(5,150)=8.65$, $p<.001$, $\eta_p^2=.224$). This position effect corresponds to a recency effect, as illustrated by Figure S3, which shows larger weights for the last numbers of the sequence, that are the most recent at the time of the report.

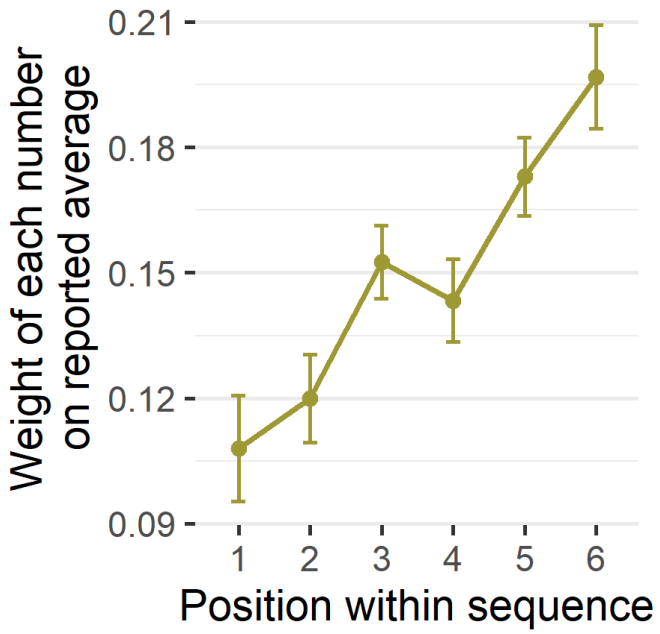

Figure S3. Weight of each number on the reported average, in a linear regression by position of the number in the sequence. Error bars represent mean and s.e.m. across participants.

4. Experiment 3. The variability effect does not come from an automatic boost on local confidence

A possible explanation for the variability effect on local confidence would be that individuals automatically increase their local confidence when presented with heterogeneous trials in terms of difficulty. If this was the case, we should find that the variability effect is more pronounced for latter trials in the set since the variability in the set becomes more and more apparent as trials pass. To test this hypothesis, we ran an ANOVA on the variability effect on local confidence (local confidence in high variability minus local confidence in low variability) by the position of the trial in the set (between 1 and 6) as within participant factors. This analysis indicated no effect of the position of the trial in the set ($F(5,150)=.449$, $p=.814$, $\eta_p^2=.015$). Pairwise comparisons between positions were all not significant ($p>.05$). As shown in Figure S4, the variability effect on local confidence does not increase as trials pass: if anything, the effect seems to become larger with the position of the trial which would go in the opposite direction that predicted by the hypothesis.

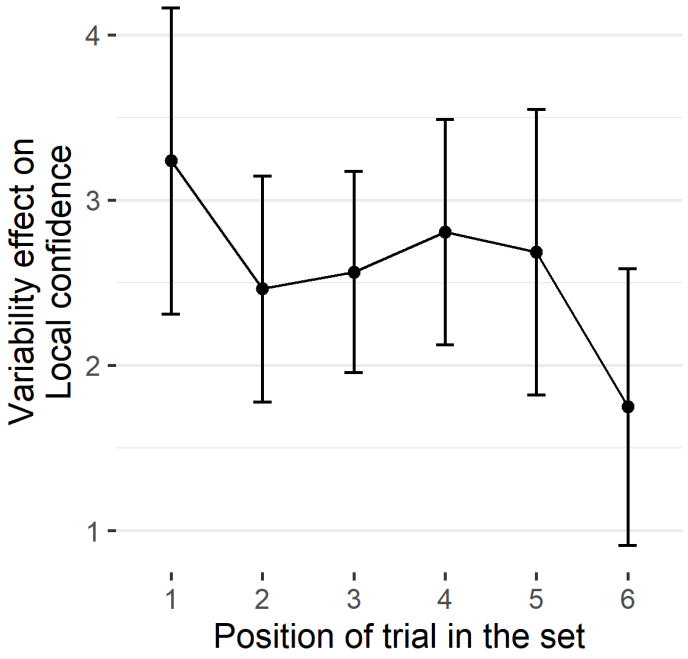

Figure S4. Variability effect on local confidence by position of the trial in the set. Error bars represent mean and s.e.m. across participants.

5. Experiment 3. Rationalizing the variability effect

As detailed below, a possible explanation for the variability effect on confidence is that it is a consequence of a distortion of objective probabilities at the level of local confidence. Moreover, we argue that such a distortion of probabilities could be rationalized by assuming individuals misperceive the prior probability over stimuli' s difficulty.

First, to understand how probabilistic distortion could generate the variability effect, we consider a framework similar to that of Experiment 3. For simplicity, we assume that the observer faces a 2 alternative forced-choice task (instead of a 4-afc in Experiment 3). He has to categorize whether a stimulus S corresponds to one category (S>0) or the other (S<0). Furthermore, we assume similar variability conditions and difficulty levels as in Experiment 3. The observer can either face a high variability set, with equal proportions of hard trials calibrated at 55% of correct responses ($S = x_{55}$ if S>0 and $S = x_{45}$ if S<0) and easy trials calibrated at 95% of correct responses ($S = x_{95}$ if S>0 and $S = x_{05}$ if S<0), or she can face a low variability set with only intermediate trials calibrated at 75% of correct responses ($S = x_{75}$ if S>0 and $S = x_{25}$ if S<0). For both types of sets, the expected probability of being correct is thus equal to 75% in both sets.

Suppose now that individuals distort objective probabilities of being correct according to a linear transformation in log-odds as suggested in Zhang and Maloney (2012). This distortion is described in eq. 1, with p_{subj} the subjective probability and p_{obj} the objective probability of being correct.

$$\log\left(\frac{p_{subj}}{1-p_{subj}}\right) = \gamma \times \log\left(\frac{p_{obj}}{1-p_{obj}}\right) + (1-\gamma) \times \log\left(\frac{p_0}{1-p_0}\right) \qquad \text{(eq.1)}$$

Parameters p_0 and γ correspond to the fixed point and to the slope of the distortion curve, illustrated in Figure S5. Here, we take the values $p_0 = .75$ and $\gamma = .5$ for these parameters, but the present demonstration would work under any convex distortion.

By definition of parameter p_0, the observer estimates correctly the probability to be correct of intermediate trials and is thus well calibrated in low variability sets. In contrast, he overestimates the probability to be correct of hard trials by 10.69 points (confidence= 65.69% vs actual probability to be correct = 55%) and underestimates the probability to be correct of easy trials by 6.70 points (confidence=88.30 vs actual probability to be correct=95%). Thus, in high variability sets, he shows an overestimation on average across trials (as illustrated by the blue dot in Figure S3). This replicates our variability effect since average overconfidence will then be higher in high variability sets than in low variability sets (by 2 points).

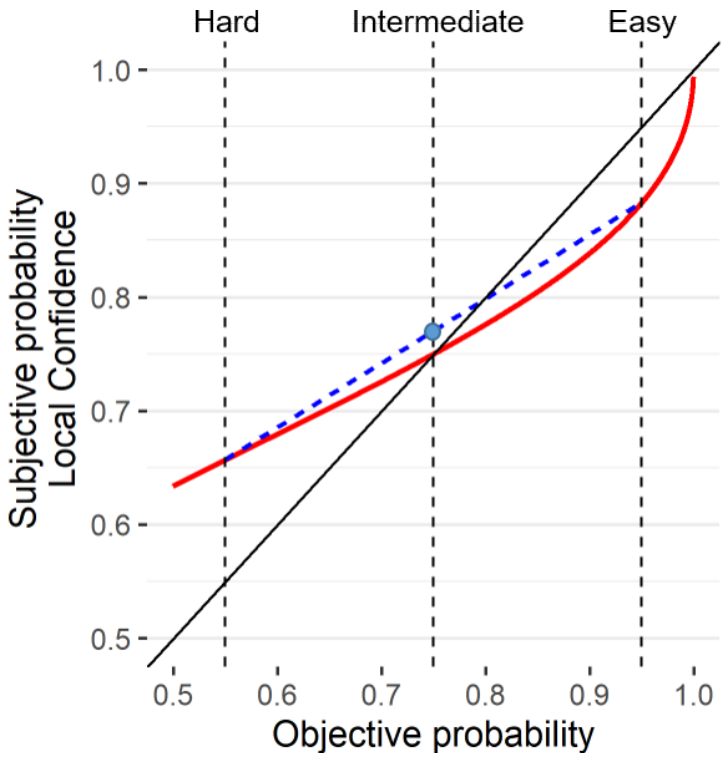

Figure S5. Subjective evaluation of probabilities against objective probabilities. The red curve represents a linear transformation in log odds of objective probabilities as in Zhang and Maloney (2012) with $p_0 = .75$ and $\gamma = .5$. From left to right, vertical dashed lines represent objective probability of .55 (hard trials), .75 (intermediate trials) and .95 (easy trials) respectively. The blue dashed line connects subjective evaluation of objective probability .55 and .95.

Assuming that the distortion function is convex over the entire interval guarantees the existence of the variability effect. Interestingly, this can be derived from a SDT model of confidence, assuming individuals do not properly estimate the prior distribution of stimuli difficulty.

To do so, assume that when presented with stimulus S, because of internal noise, the observer receives a signal (noted s) normally distributed around the true stimulus value S, with standard deviation 1, without loss of generality. Consider an observer who believes that all sets are low variability sets meaning that all trials in all sets are intermediate trials, calibrated at 75%.

Suppose that the observer answered “S>0” to a trial following a signal s. Observer’s confidence will be:

$P(S > 0|s) = \frac{P(s|S = x_{75})}{P(s|S = x_{25}) + P(s|S = x_{75})} = \frac{h(x_{75})}{h(x_{25}) + h(x_{75})}$ in all sets, with h the density of a normal distribution with mean s and variance 1.

Although confidence is correctly computed in low variability sets (because the prior is correct for those sets), it is incorrect in the case of high variability sets, due to the wrong prior assumption. In those sets, confidence should be:

$$P(S > 0|s) = \frac{P(s|S = x_{95}) + P(s|S = x_{55})}{P(s|S = x_{95}) + P(s|S = x_{55}) + P(s|S = x_{05}) + P(s|S = x_{45})} = \frac{h(x_{95}) + h(x_{55})}{h(x_{95}) + h(x_{55}) + h(x_{05}) + h(x_{45})}$$

Panel A of Figure S6 plots, for high variability sets, the observer’s confidence in red and this “correct” confidence in blue (that is the confidence computed under a prior of high variability). For most values of s (on the x-axis), the observer’s confidence exceeds the “correct” confidence, resulting in overconfidence overall for high variability sets. By contrast, for low variability sets, the prior is correct and the observer is well-calibrated.

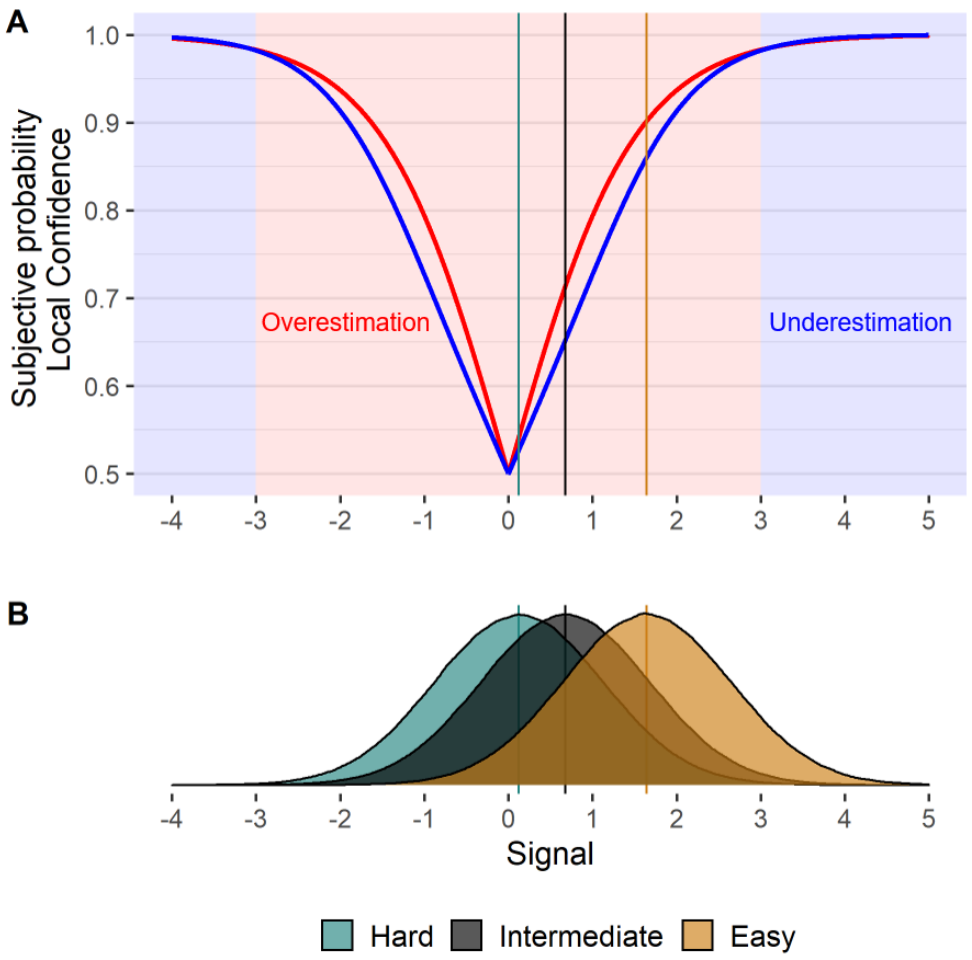

Figure S6. (A) Confidence as a function of the signal which is received. The blue curve represents the "correct" confidence, under a high variability prior. The red curve represents the observer's confidence, under a low variability prior. The red area represents signal values for which the red curve is above the blue curve and the blue area the opposite. (B) The green, black and yellow density represents the distribution of signals for hard, intermediate and easy trials, respectively. The median signal value for hard, intermediate and easy trials are represented by a green, black and yellow line, respectively.