SUPPLEMENTARY INFORMATION

Confidence matching in group decision-making

Dan Bang*, Laurence Aitchison, Rani Moran, Santiago Herce Castañón, Banafshe Rafiee, Ali Mahmoodi, Jennifer Y. F. Lau, Peter E. Latham, Bahador Bahrami, and Christopher Summerfield

*Correspondence to: danbang.db@gmail.com

SUPPLEMENTARY METHODS

Computational model

We developed a simple signal detection model (to understand how joint accuracy (fraction of correct joint decisions) varies with differences between group members' expertise and mean confidence and to establish an optimal benchmark against which empirical group performance could be compared (**Figure 3**). The model assumes that three cognitive processes govern confidence: (i) the agent receives noisy sensory evidence, (ii) the agent computes an internal estimate of the evidence strength and (iii) the agent maps the internal estimate onto a response (decision and confidence) by applying a set of thresholds. The level of sensory noise determines the agent's expertise and the set of thresholds determines the agent's mean confidence.

Model description

We assumed that on each trial an agent receives noisy sensory evidence, x, sampled from a Gaussian distribution, $x \in N(s, \sigma^2)$. The mean, s, is the difference in contrast between the second and the first display at the target location. As such, s is drawn uniformly from the set $s \in S = \{-.15, -.07, -.035, -.015, .015, .035, .07, .15\}$ – the sign of s indicates the target display (negative: 1st; positive: 2nd) and its absolute value indicates the contrast added to the target. The standard deviation, σ, describes the level of sensory noise and is the same for all stimuli.

We modelled the internal estimate of the evidence strength as the raw sensory evidence, $z = x$. The internal estimate thus ran from large negative values, indicating a high probability that the target was in the first display, through values near 0, indicating high uncertainty, to large positive values, indicating a high probability that the target was in the second display. We chose this formulation for mathematical simplicity but note that our analysis would show the same results for any model in which the internal estimate is a monotonic function of the sensory evidence, including probabilistic estimates such as $z = P(s > 0|x)$.

We assumed that the agent maps the internal estimate onto a response, r, by applying a set of thresholds, $\boldsymbol{\theta}$. As in our experiments, the responses range from -6 to -1 and 1 to 6 – the sign of r indicates the decision (negative: 1st; positive: 2nd) and its absolute value indicates the confidence ($c = |r|$). The thresholds, $\boldsymbol{\theta}$, determine the probability distribution over responses

$$p_i \equiv P(r = i) = \begin{cases} P(z \le \theta_{-6}) & i = -6 \\ P(\theta_{i-1} < z \le \theta_i) & -6 < i \le -1, \quad 2 \le i < 6 \\ P(\theta_{-1} < z \le \theta_1) & i = 1 \\ P(z > \theta_5) & i = 6 \end{cases}.$$

There is no criterion θ_6, because $r = 6$ corresponds to z exceeding θ_5. There is a one-to-one relationship between thresholds and probabilities, so it is easy to find the thresholds corresponding to a given response distribution (as we will do shortly).

Deriving accuracy for an agent

We calculated the accuracy (fraction of correct decisions) of an agent, given a level of sensory noise, σ, and a response distribution, p_i, where $i = \pm 1,2, \dots, 6$, as follows.

We first calculated the thresholds, $\boldsymbol{\theta}$, that produced the response distribution p_i over the entire stimulus set $S = \{-.15, -.07, -.035, -.015, .015, .035, .07, .15\}$. In particular, we found (using MATLAB's 'fzero' function) thresholds θ_i, where $i = -6, -5, \dots, -1,1,2, \dots 5$, such that

$$\sum\nolimits_{j \le i} p_j = \frac{1}{8} \sum\nolimits_{s \in S} \Phi\left(\frac{\theta_i - s}{\sigma}\right)$$

where Φ is the Gaussian cumulative density fucntion. We then calculated for each stimulus, $s \in S$, the predicted response distribution, denoted $p_{i,s}$,

$$p_{i,s} = \begin{cases} \Phi\left(\frac{\theta_{-6} - s}{\sigma}\right) & i = -6 \\ \Phi\left(\frac{\theta_i - s}{\sigma}\right) - \Phi\left(\frac{\theta_{i-1} - s}{\sigma}\right) & -6 < i < 6 \\ 1 - \Phi\left(\frac{\theta_5 - s}{\sigma}\right) & i = 6 \end{cases}.$$

Thus, the accuracy of an agent was given by

$$a_{\text{agent}} = \frac{\sum_{s \in S, s>0} \sum_{i=1}^{6} p_{i,s} + \sum_{s \in S, s<0} \sum_{i=-6}^{-1} p_{i,s}}{8}$$

Deriving joint accuracy for a pair of agents

We calculated the joint accuracy (fraction of correct joint decisions) for a pair of agents as follows:

We computed the predicted response distribution for each stimulus $s \in S$ for group member 1, $p_{i,s,1}$, and group member 2, $p_{i,s,2}$. We then calculated for each stimulus $s \in S$ the probability that group member 1 makes response i_1 and group member 2 makes response i_2 as $\tilde{p}_{i_1,i_2,s} = p_{i,s,1} p_{i,s,2}$. The response combinations that yield a correct response for a positive-mean Gaussian distribution are those for which $i_1 + i_2 > 0$. The response combinations that yield a correct response for a negative-mean Gaussian distribution are those for which $i_1 + i_2 < 0$. Additionally, response combinations for which i_1+i_2 = 0 – that is, confidence ties – yield a correct response with probability .5.

The join accuracy of a pair of agents is thus given by

$$a_{\text{joint}} = \frac{1}{8}\left[\sum_{s \in S, s>0}\left(\sum_{i_1+i_2>0} \tilde{p}_{i_1,i_2,s} + \frac{1}{2}\sum_{i_1+i_2=0} \tilde{p}_{i_1,i_2,s}\right) + \sum_{s \in S, s<0}\left(\sum_{i_1+i_2<0} \tilde{p}_{i_1,i_2,s} + \frac{1}{2}\sum_{i_1+i_2=0} \tilde{p}_{i_1,i_2,s}\right)\right]$$

CONFIDENCE LANDSCAPES

Calculating a set of maximum entropy distributions

For a given mean there are, obviously, many possible distributions. Here we choose the one that maximized the entropy, denoted H,

$$H \equiv -\sum_{i=1}^{6} p_i \log p_i$$

The maximum entropy distribution is found using Lagrange multipliers,

$$\frac{\partial}{\partial p_i}\left[H + \lambda_1\left(\sum_{i=1}^{6} p_i - 1\right) + \lambda_2\left(\sum_{i=1}^{6} i p_i - c\right)\right] = 0$$

$$\frac{\partial}{\partial \lambda_1}\left[H + \lambda_1\left(\sum_{i=1}^{6} p_i - 1\right) + \lambda_2\left(\sum_{i=1}^{6} i p_i - c\right)\right] = 0$$

$$\frac{\partial}{\partial \lambda_2}\left[H + \lambda_1\left(\sum_{i=1}^{6} p_i - 1\right) + \lambda_2\left(\sum_{i=1}^{6} i p_i - c\right)\right] = 0$$

where c is the mean. Solving these equations gives the maximum entropy distribution, subject to the constraints that the distribution sums to 1, $\sum_{i=1}^{6} p_i = 1$, and has the right mean confidence, c, $\sum_{i=1}^{6} ip_i = c$ and $p_i \geq 0$. For $c = 1$ the unique distribution that satisfies these constraints is $\underline{p} = (1,0,0,0,0,0)$. Similarly, for $c = 6$, $\underline{p} = (0,0,0,0,0,1)$ is the solution. For any intermediate value for the mean, $1 \leq c \leq 6$, the solution is

$$p_i = \frac{e^{i\lambda_2}}{\sum_{j=1}^{6} e^{j\lambda_2}}$$

with λ_2 chosen by solving the constraint

$$c = \frac{\sum_{j=1}^{6} j e^{j\lambda_2}}{\sum_{j=1}^{6} e^{j\lambda_2}}$$

which can be done with MATLAB's "fzero" function. We transformed confidence distributions (1 to 6) to response distributions (-6 to -1 and 1 to 6) by assuming symmetry around 0.

SUPPLEMENTARY NOTES

Comparing observed to optimal joint accuracy

In **Figure 3D**, we show that the ratio of the observed joint accuracy to that expected under the optimal solution, $a_{\mathrm{emp}}/a_{\mathrm{opt}}$, approaches 1 as the ratio of the accuracy of the less accurate group member to that of the more accurate group member, $a_{\mathrm{min}}/a_{\mathrm{max}}$, approaches 1 – a pattern which is expected under confidence matching. This analysis is, however, subject to potential confounds. Our aim here is to explain what these confounds are and show that each of them can be ruled out.

First, we might find a divergence between the observed joint accuracy and that expected under the optimal solution because the only source of response variation in our model is sensory noise; unlike participants, the model never makes response errors or has lapses of attention. To correct for this potential mismatch, we re-defined joint accuracy under the optimal solution as:

$$\tilde{a}_{\mathrm{opt}} = a_{\mathrm{opt}} - (a_{\mathrm{fit}} - a_{\mathrm{emp}})$$

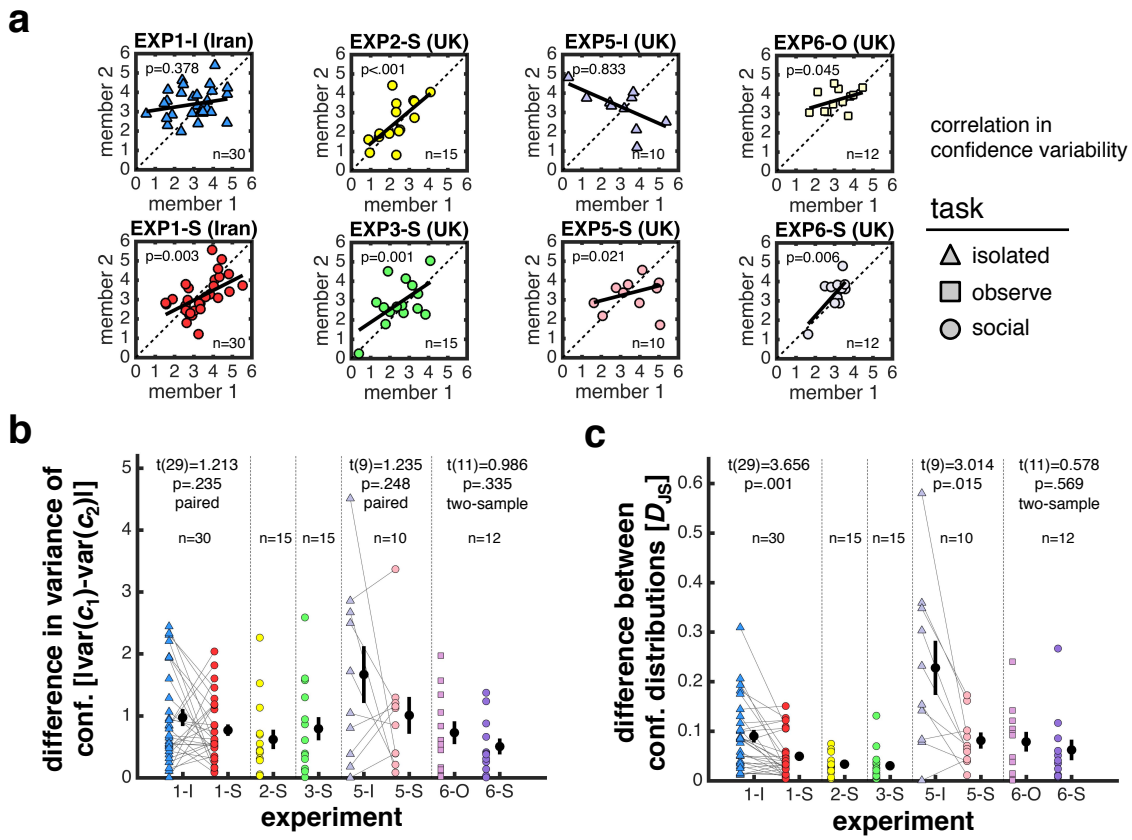
where a_{opt} is the maximum of a given confidence landscape as before and a_{fit} is the joint accuracy derived under group members' observed response distributions. The additional term, $(a_{\mathrm{fit}} - a_{\mathrm{emp}})$, corrects for any 'overestimation' of joint accuracy introduced by our modelling procedure. Critically, when re-defining optimality as $a_{\mathrm{emp}}/\tilde{a}_{\mathrm{opt}}$, we still observe a positive correlation between optimality and the similarity of group members' accuracy ($r(80) = .682$, $p < .001$, Pearson).

This analysis, however, may still be subject to two additional confounds. First, $\tilde{a}_{\mathrm{opt}}$ might diverge from a_{emp} simply because $\tilde{a}_{\mathrm{opt}}$ was derived under maximum entropy distributions, which – for any given mean confidence – usually reduces the number of confidence ties. Second, the level of optimality itself, $a_{\mathrm{emp}}/\tilde{a}_{\mathrm{opt}}$, might be smaller for similar group members, $a_{\mathrm{min}}/a_{\mathrm{max}} \approx 1$, than dissimilar group members, $a_{\mathrm{min}}/a_{\mathrm{max}} \ll 1$, due to range effects. For the most dissimilar group members, the difference between the minimum and the maximum value of a given confidence landscape is about .1. In contrast, for the most similar group members, the difference between the minimum and the maximum value of a given confidence landscape is only about .05. Thus, there was less room for the latter to deviate from the joint accuracy expected under the optimal solution.

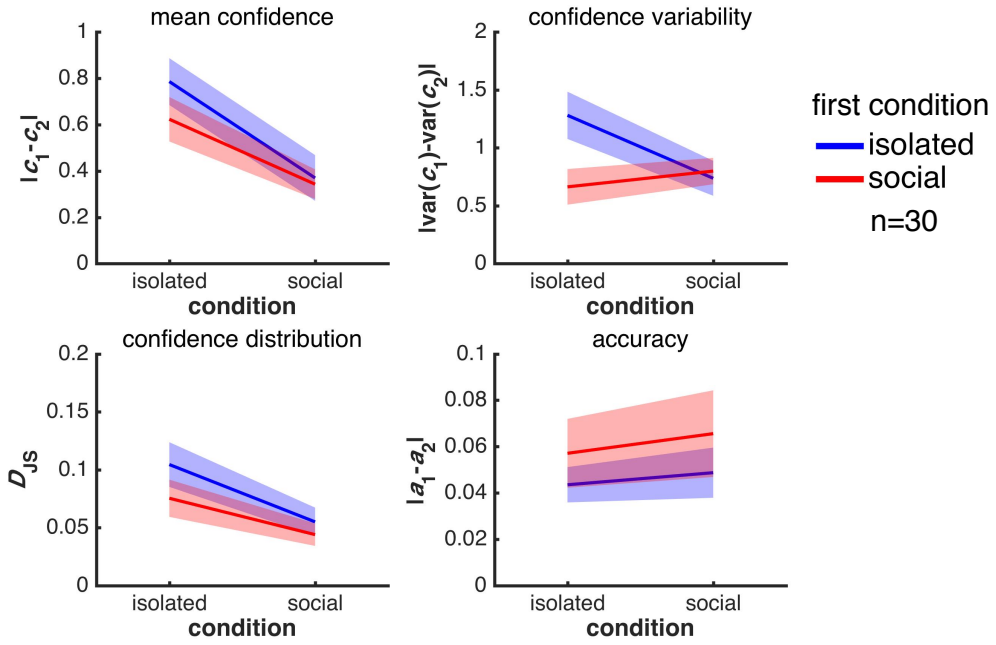
We therefore repeated the above analysis but now re-defining optimality as:

$$a_{\mathrm{emp}}/\tilde{a}_{\mathrm{opt}} = (a_{\mathrm{maxent}} - a_{\mathrm{anti}})/(a_{\mathrm{opt}} - a_{\mathrm{anti}})$$

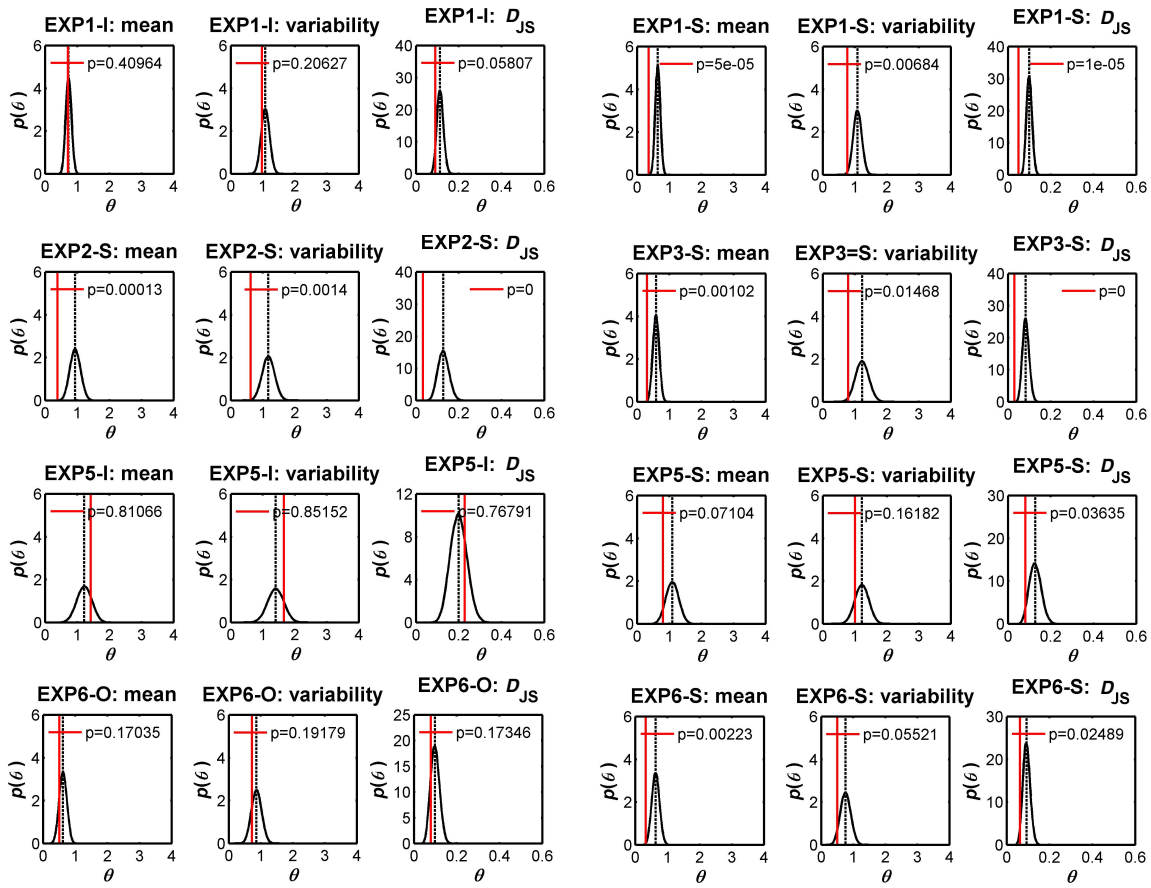
where a_{anti} is the minimum value of a confidence landscape and a_{maxent} is the value of the landscape coordinate indexed by group member's observed mean confidence. First, by using a_{maxent} instead of a_{emp} in the numerator, we control not only for model misspecification but also for the use of maximum entropy distributions. Second, the subtraction of a_{anti} in both the numerator and denominator controls for range effects. Importantly, in line with the above results, we still find a positive correlation between optimality and the similarity of group members' accuracy ($r(80) = .469$, $p < .001$, Pearson).



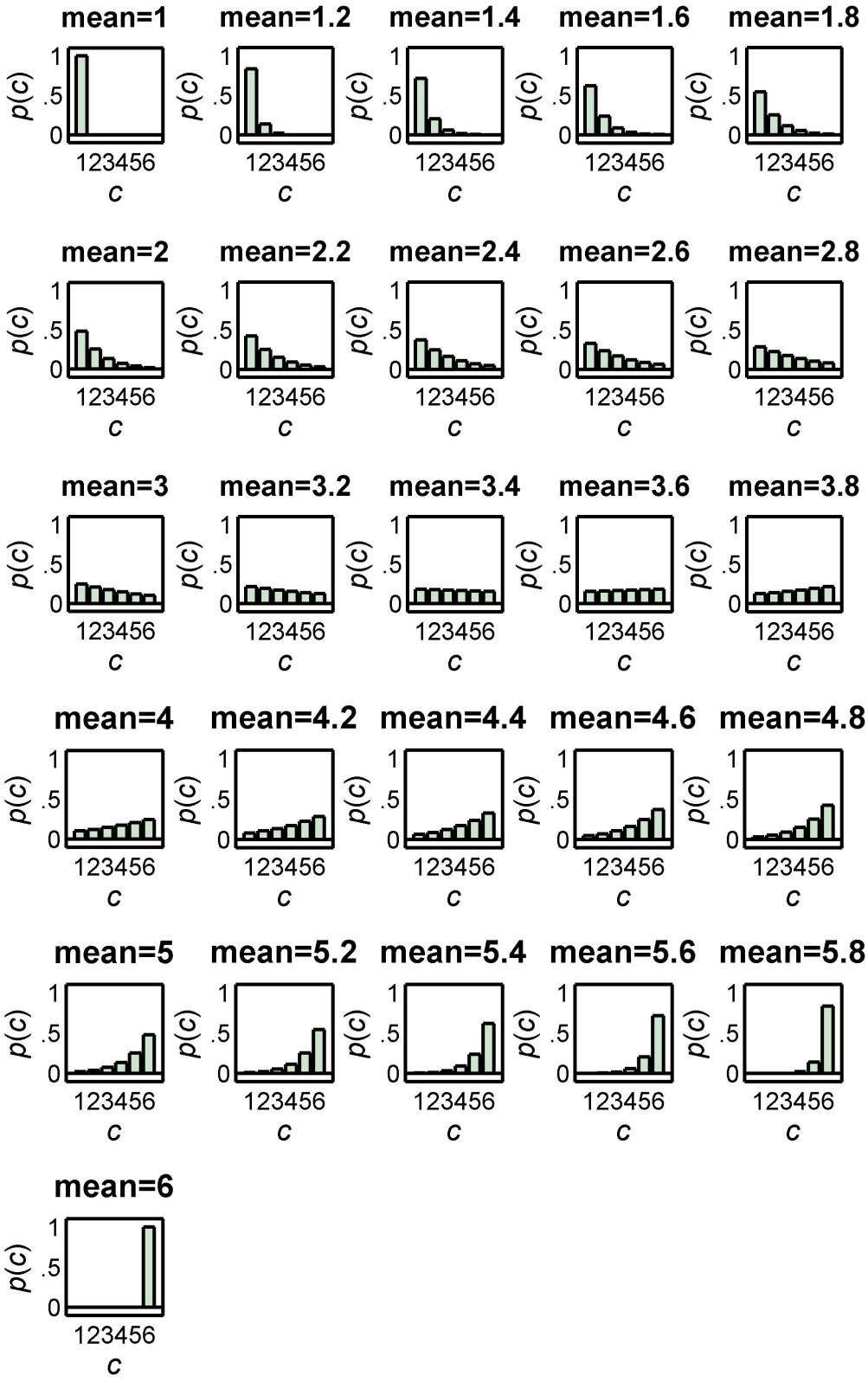
Supplementary Figure 1 | Evidence for matching of confidence variability and confidence distributions. a, Correlation in confidence variability. The axes show the variance of group members' confidence, $\mathrm{var}(c_1)$ and $\mathrm{var}(c_2)$. The results indicate that group members adapted to each other's confidence variability in the social task only. Each dot is a group. Each line is the best-fitting line of a robust regression; because the sorting of group members into 1 and 2 is arbitrary, we show the *p*-value for the slope of the best-fitting line averaged across 10^5 separate regressions, for each randomly re-labelling the members of a group as 1 and 2. **b**, Convergence in confidence variability. The y-axis shows the absolute difference between the variance of group members' confidence, $|\mathrm{var}(c_1) - \mathrm{var}(c_2)|$. The results are not conclusive; in EXP1, EXP5 and EXP6, there are only trends towards convergence in confidence variability in the social task compared to the other tasks. Each black dot is data averaged across groups. Each coloured dot is a group; the lines connect group data when the same pairing of group members was used in two conditions. Error bars are 1 SEM. **c**, Convergence in confidence distributions. The y-axis shows the Jensen-Shannon divergence, D_{JS}, between group members' confidence distributions. The results provide strong support for convergence in confidence distributions in the social task compared to the other tasks, except in the case of EXP6. In EXP3, the continuous confidence values were discretised to the range 1 to 6 in steps of 1. Each black dot is data averaged across groups. Each coloured dot is a group; the lines connect group data when the same pairing of group members was used in two conditions. Error bars are 1 SEM. See **Supplementary Figure 3** for complementary permutation-based tests.



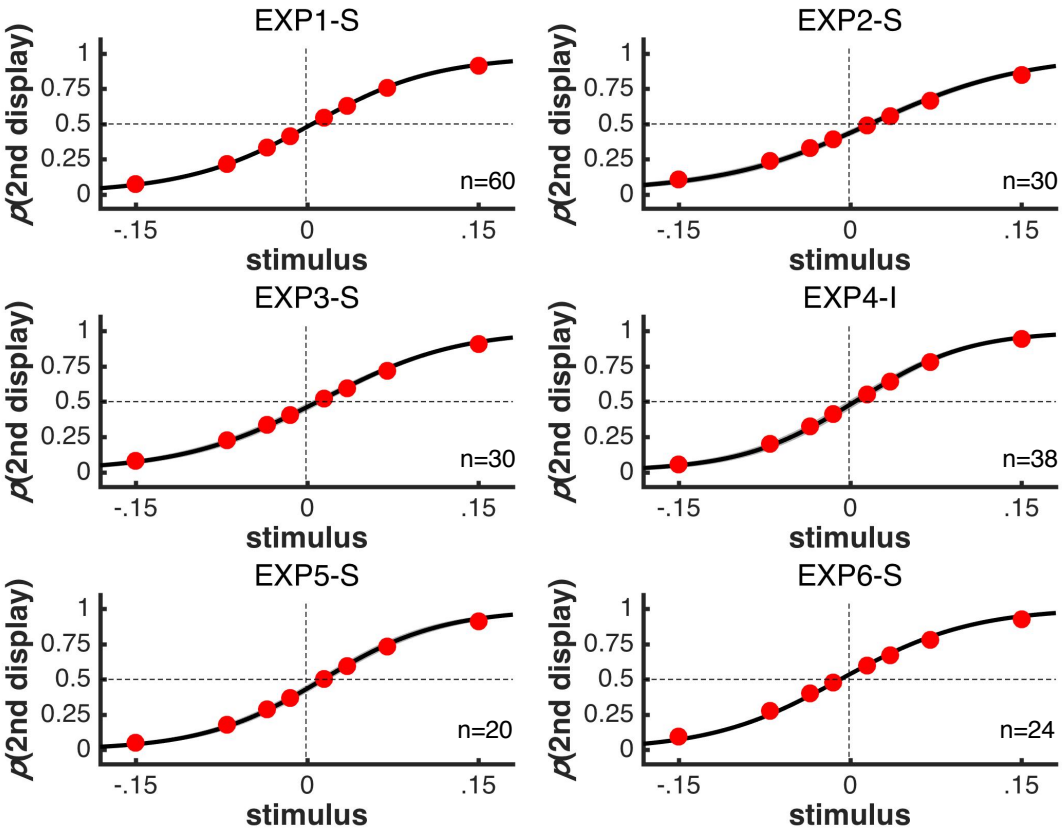
Supplementary Figure 2 | Results from Experiment 1 split by condition order. The x-axis shows the current condition, the y-axis shows the measure of interest and the colour denotes whether groups first performed the isolated condition (blue) or the social condition (red). We tested the effect of condition order using an ANOVA with the measure of interest as within-subject factor (isolated versus social) and the condition order as between-subject factor (isolated first versus social first). The interaction between the two factors was only significant for the differences between the variance of group members' confidence (variance: $F_{1,28} = 4.160$, $p = .041$; all others: $F_{1,28} < 0.700$, $p > .400$. D_{JS}: Jensen-Shannon divergence.



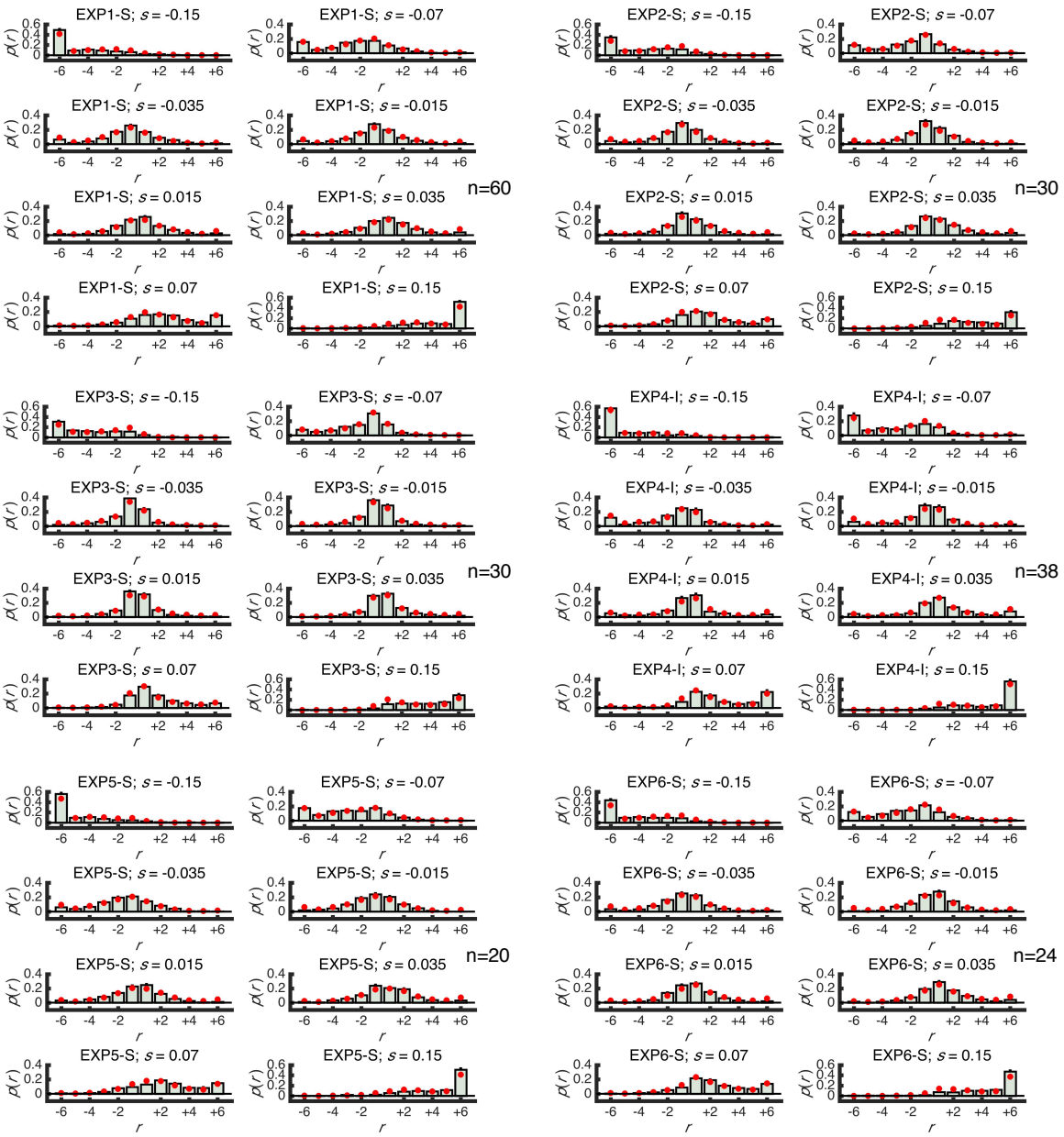
Supplementary Figure 3 | Null distributions for permutation testing. We created for each measure of interest, ϑ, a distribution under the null hypothesis, $p(\vartheta)$, by randomly repairing participants and computing the mean of measured values for each set of repaired participants (we simulated 10^6 sets in total). By asking whether the observed mean value (red line) for a given measure was smaller than 95% of the values from the null distribution (i.e., $p < .05$, one-tailed), we could test whether the observed mean value was specific to the true pairing of group members – we would expect this to be the case if it was the result of dynamic interaction between group members. The permutation-based approach unequivocally shows that the mean values observed in the social task are specific to the true pairing of group members. The plots show the probability density function for each measure of interest. The dotted line shows the mean of each null distribution. The *p*-value indicates the proportion of values from the null distribution that were smaller than the observed mean value. Mean: $|c_1 - c_2|$. Variability: $|\mathrm{var}(c_1) - \mathrm{var}(c_2)|$. D_{JS}: Jensen-Shannon divergence.



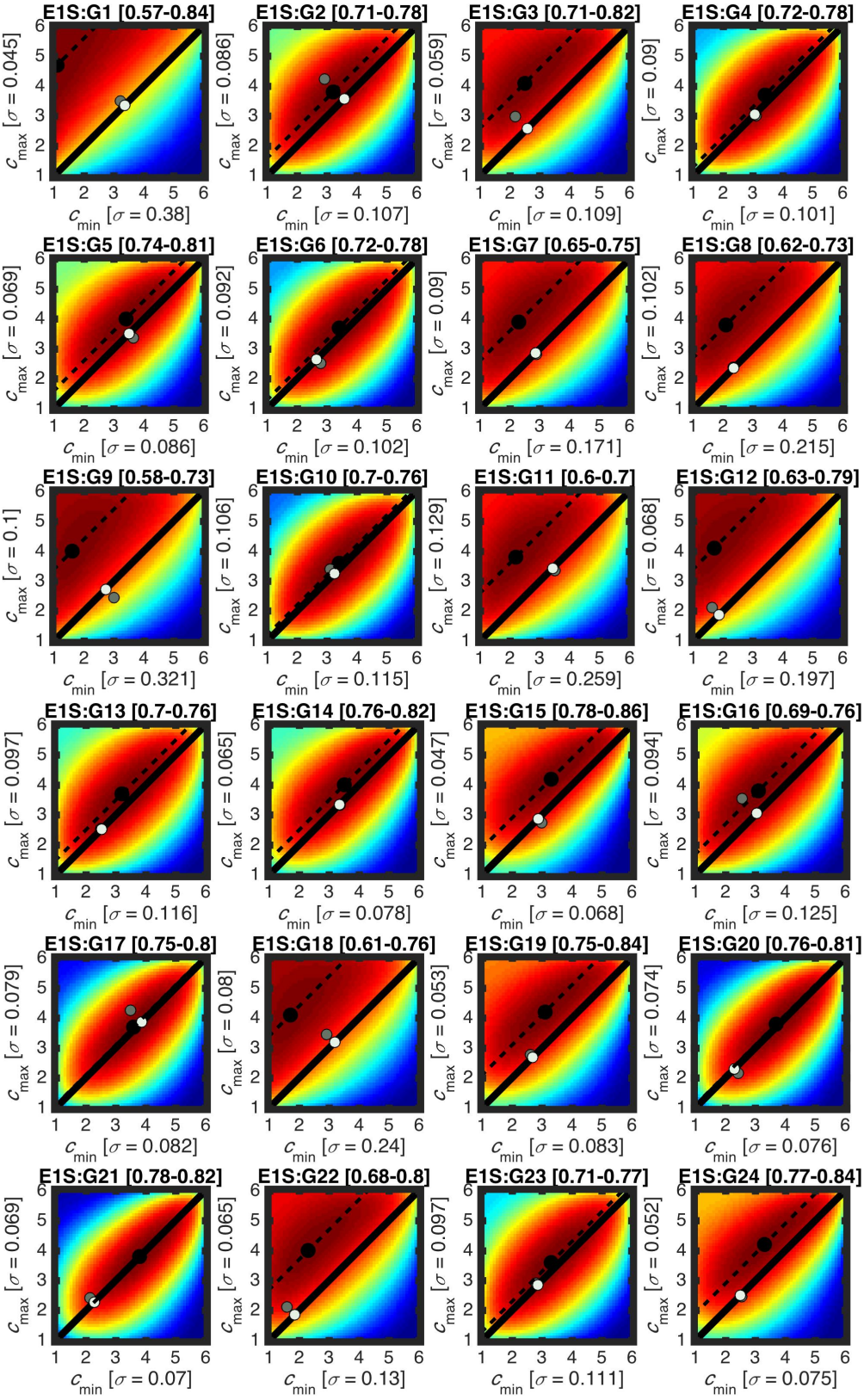
Supplementary Figure 4 | Maximum entropy distributions. The x-axis shows the confidence levels. The y-axis shows the proportion of times that each confidence level is selected. We transformed confidence distributions (1 to 6) to response distributions (-6 to -1 and 1 to 6) by assuming symmetry around 0. We only display a subset of the maximum entropy distributions (full set: 1 to 6 in steps of .1).

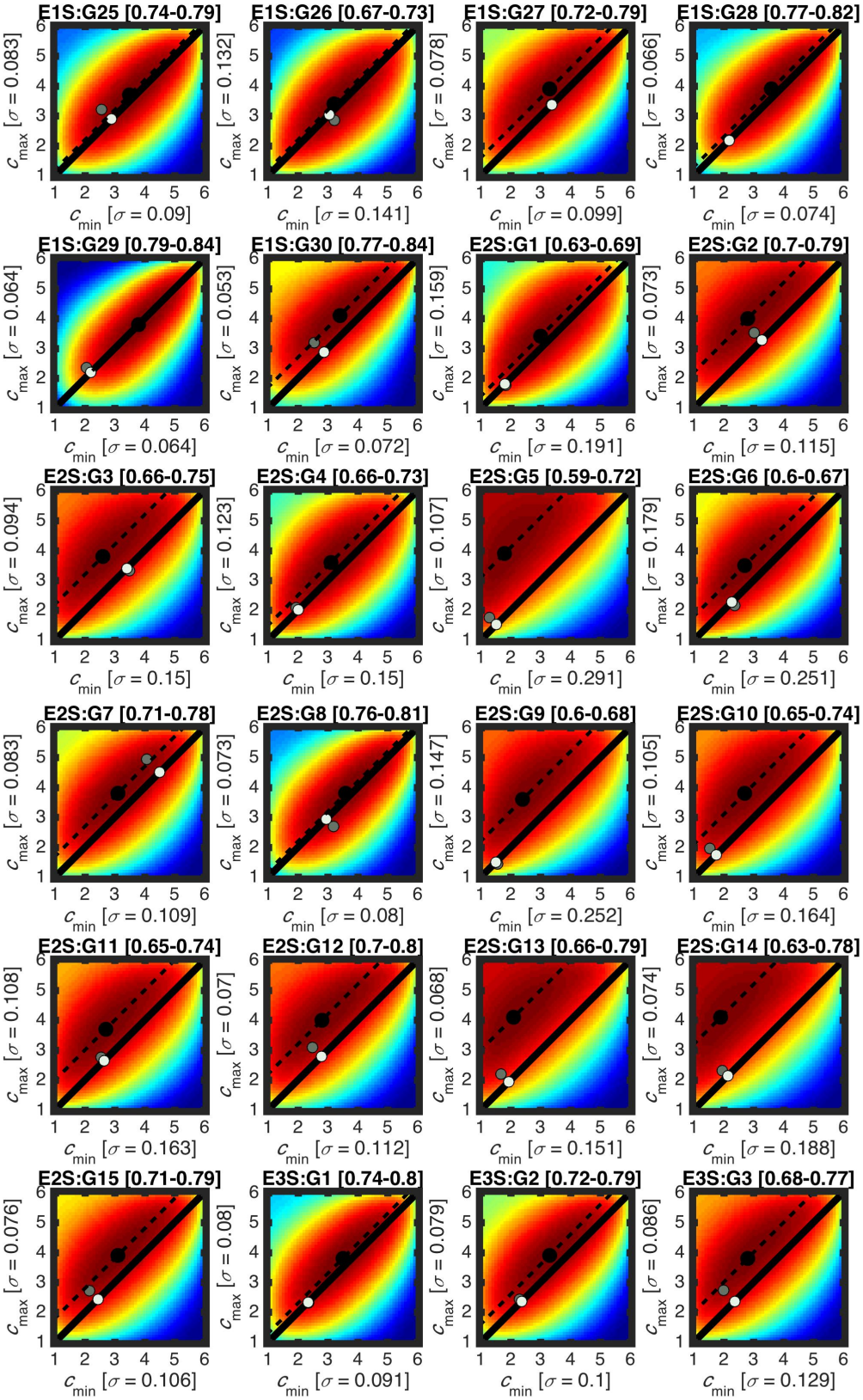


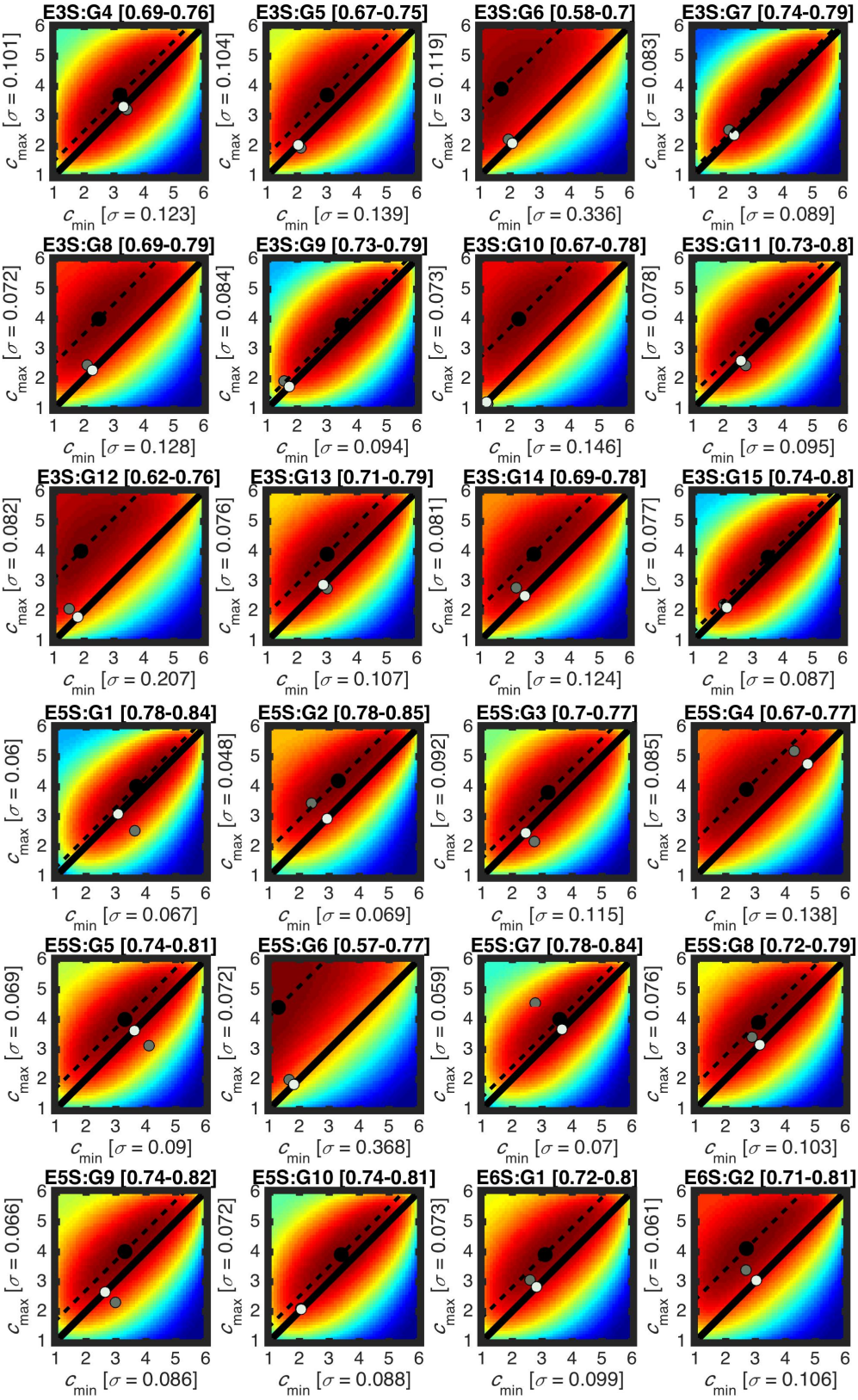
Supplementary Figure 5 | Comparison between empirical and model psychometric functions. The x-axis shows the stimulus. The y-axis shows the proportion of times that the second display was selected for a given value of the stimulus. The plots were created by fitting psychometric functions to the empirical and the model data. We computed R^2-values across participants (mean ± SD) to evaluate the model fits: EXP1-S: R^2 = .825 ± .358; EXP2: R^2 = .892 ± .089; EXP3: R^2 = .947 ± .030; EXP4-I: R^2 = .844 ± .135; EXP5-S: R^2 = .823 ± .405; EXP6-S: R^2 = .862 ± .193. Black line: empirical psychometric functions averaged across participants. Red dots: model psychometric functions (best fits) averaged across participants; dots only shown for the stimuli used in the experiments. Shaded area/error bars are 1 SEM (not visible in these plots).

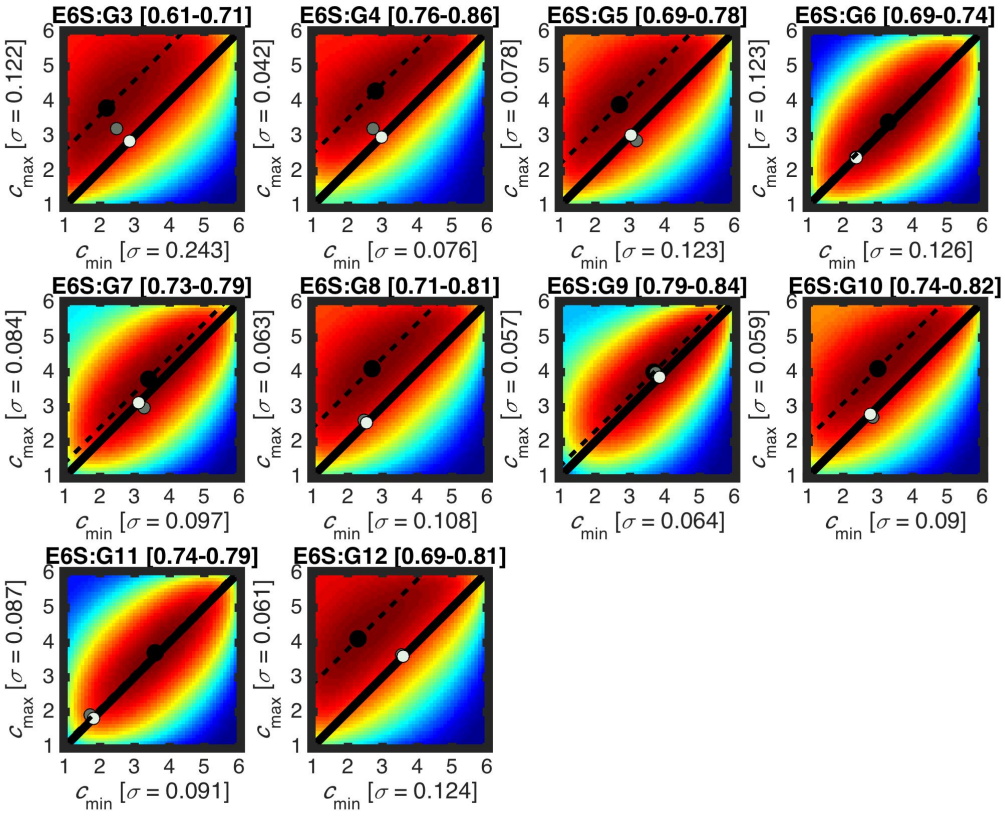


Supplementary Figure 6 | Comparison between empirical and model response distributions. The x-axis shows the response values (from -6 to 1 and 1 to 6). The y-axis shows the proportion of times that each response was made. The title of each subplot indicates the experiment and the stimulus. We computed R^2-values across participants (*mean* ± *SD*) to evaluate the model fits. EXP1-S: R^2 = .606 ± .354; EXP2: R^2 = .777 ± .136; EXP3: R^2 = .838 ± .078; EXP4-I: R^2 = .613 ± .251; EXP5-S: R^2 = .761 ± .146; EXP6-S: R^2 = .635 ± .206. The reason why the empirical and the model response distributions are not identical is that the model was fitting using the response distribution observed across all stimuli. Grey: empirical response distributions averaged across participants. Red: model response distributions (best fits) averaged across participants. Error bars are 1 SEM.

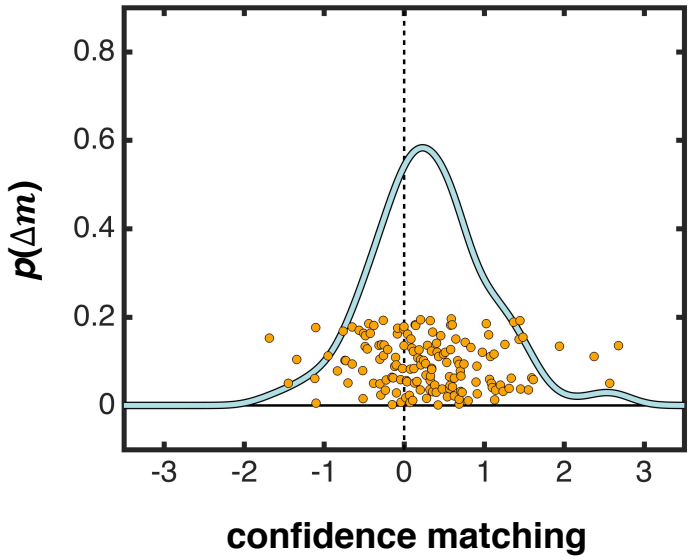

E1S:G1 [0.57-0.84]
E1S:G2 [0.71-0.78]
E1S:G3 [0.71-0.82]
E1S:G4 [0.72-0.78]
E1S:G5 [0.74-0.81]
E1S:G6 [0.72-0.78]
E1S:G7 [0.65-0.75]
E1S:G8 [0.62-0.73]
E1S:G9 [0.58-0.73]
E1S:G10 [0.7-0.76]
E1S:G11 [0.6-0.7]
E1S:G12 [0.63-0.79]
E1S:G13 [0.7-0.76]
E1S:G14 [0.76-0.82]
E1S:G15 [0.78-0.86]
E1S:G16 [0.69-0.76]
E1S:G17 [0.75-0.8]
E1S:G18 [0.61-0.76]
E1S:G19 [0.75-0.84]
E1S:G20 [0.76-0.81]
E1S:G21 [0.78-0.82]
E1S:G22 [0.68-0.8]
E1S:G23 [0.71-0.77]
E1S:G24 [0.77-0.84]
c_max [σ = 0.045]
c_min [σ = 0.38]
c_max [σ = 0.086]
c_min [σ = 0.107]
c_max [σ = 0.059]
c_min [σ = 0.109]
c_max [σ = 0.09]
c_min [σ = 0.101]
c_max [σ = 0.069]
c_min [σ = 0.086]
c_max [σ = 0.092]
c_min [σ = 0.102]
c_max [σ = 0.09]
c_min [σ = 0.171]
c_max [σ = 0.102]
c_min [σ = 0.215]
c_max [σ = 0.1]
c_min [σ = 0.321]
c_max [σ = 0.106]
c_min [σ = 0.115]
c_max [σ = 0.129]
c_min [σ = 0.259]
c_max [σ = 0.068]
c_min [σ = 0.197]
c_max [σ = 0.097]
c_min [σ = 0.116]
c_max [σ = 0.065]
c_min [σ = 0.078]
c_max [σ = 0.047]
c_min [σ = 0.068]
c_max [σ = 0.094]
c_min [σ = 0.125]
c_max [σ = 0.079]
c_min [σ = 0.082]
c_max [σ = 0.08]
c_min [σ = 0.24]
c_max [σ = 0.053]
c_min [σ = 0.083]
c_max [σ = 0.074]
c_min [σ = 0.076]
c_max [σ = 0.069]
c_min [σ = 0.07]
c_max [σ = 0.065]
c_min [σ = 0.13]
c_max [σ = 0.097]
c_min [σ = 0.111]
c_max [σ = 0.052]
c_min [σ = 0.075]


E1S:G25 [0.74-0.79]
E1S:G26 [0.67-0.73]
E1S:G27 [0.72-0.79]
E1S:G28 [0.77-0.82]
E1S:G29 [0.79-0.84]
E1S:G30 [0.77-0.84]
E2S:G1 [0.63-0.69]
E2S:G2 [0.7-0.79]
E2S:G3 [0.66-0.75]
E2S:G4 [0.66-0.73]
E2S:G5 [0.59-0.72]
E2S:G6 [0.6-0.67]
E2S:G7 [0.71-0.78]
E2S:G8 [0.76-0.81]
E2S:G9 [0.6-0.68]
E2S:G10 [0.65-0.74]
E2S:G11 [0.65-0.74]
E2S:G12 [0.7-0.8]
E2S:G13 [0.66-0.79]
E2S:G14 [0.63-0.78]
E2S:G15 [0.71-0.79]
E3S:G1 [0.74-0.8]
E3S:G2 [0.72-0.79]
E3S:G3 [0.68-0.77]
c_max [σ = 0.083]
c_min [σ = 0.09]
c_max [σ = 0.132]
c_min [σ = 0.141]
c_max [σ = 0.078]
c_min [σ = 0.099]
c_max [σ = 0.066]
c_min [σ = 0.074]
c_max [σ = 0.064]
c_min [σ = 0.064]
c_max [σ = 0.053]
c_min [σ = 0.072]
c_max [σ = 0.159]
c_min [σ = 0.191]
c_max [σ = 0.073]
c_min [σ = 0.115]
c_max [σ = 0.094]
c_min [σ = 0.15]
c_max [σ = 0.123]
c_min [σ = 0.15]
c_max [σ = 0.107]
c_min [σ = 0.291]
c_max [σ = 0.179]
c_min [σ = 0.251]
c_max [σ = 0.083]
c_min [σ = 0.109]
c_max [σ = 0.073]
c_min [σ = 0.08]
c_max [σ = 0.147]
c_min [σ = 0.252]
c_max [σ = 0.105]
c_min [σ = 0.164]
c_max [σ = 0.108]
c_min [σ = 0.163]
c_max [σ = 0.07]
c_min [σ = 0.112]
c_max [σ = 0.068]
c_min [σ = 0.151]
c_max [σ = 0.074]
c_min [σ = 0.188]
c_max [σ = 0.076]
c_min [σ = 0.106]
c_max [σ = 0.08]
c_min [σ = 0.091]
c_max [σ = 0.079]
c_min [σ = 0.1]
c_max [σ = 0.086]
c_min [σ = 0.129]

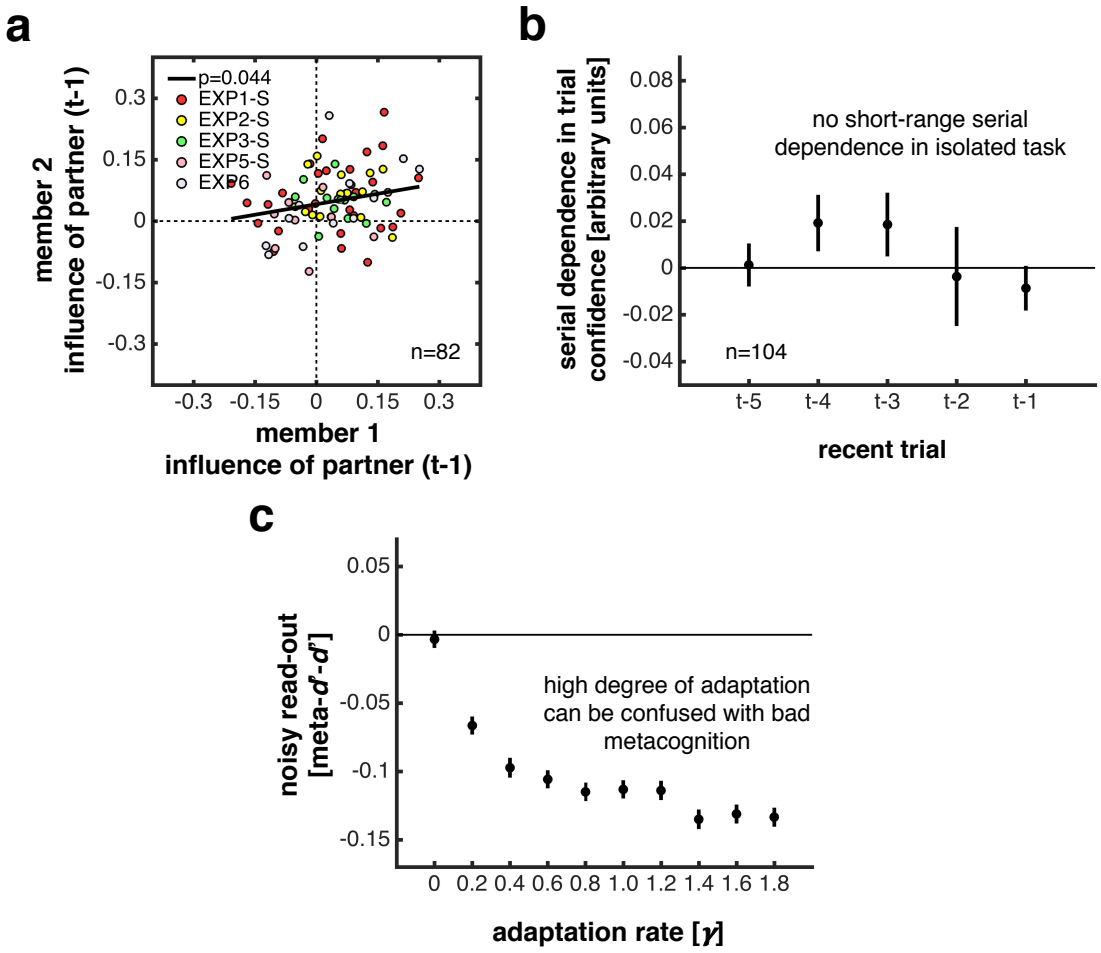

E3S:G4 [0.69-0.76]
E3S:G5 [0.67-0.75]
E3S:G6 [0.58-0.7]
E3S:G7 [0.74-0.79]
E3S:G8 [0.69-0.79]
E3S:G9 [0.73-0.79]
E3S:G10 [0.67-0.78]
E3S:G11 [0.73-0.8]
E3S:G12 [0.62-0.76]
E3S:G13 [0.71-0.79]
E3S:G14 [0.69-0.78]
E3S:G15 [0.74-0.8]
E5S:G1 [0.78-0.84]
E5S:G2 [0.78-0.85]
E5S:G3 [0.7-0.77]
E5S:G4 [0.67-0.77]
E5S:G5 [0.74-0.81]
E5S:G6 [0.57-0.77]
E5S:G7 [0.78-0.84]
E5S:G8 [0.72-0.79]
E5S:G9 [0.74-0.82]
E5S:G10 [0.74-0.81]
E6S:G1 [0.72-0.8]
E6S:G2 [0.71-0.81]
c_{max} [σ = 0.101]
c_{min} [σ = 0.123]
c_{max} [σ = 0.104]
c_{min} [σ = 0.139]
c_{max} [σ = 0.119]
c_{min} [σ = 0.336]
c_{max} [σ = 0.083]
c_{min} [σ = 0.089]
c_{max} [σ = 0.072]
c_{min} [σ = 0.128]
c_{max} [σ = 0.084]
c_{min} [σ = 0.094]
c_{max} [σ = 0.073]
c_{min} [σ = 0.146]
c_{max} [σ = 0.078]
c_{min} [σ = 0.095]
c_{max} [σ = 0.082]
c_{min} [σ = 0.207]
c_{max} [σ = 0.076]
c_{min} [σ = 0.107]
c_{max} [σ = 0.081]
c_{min} [σ = 0.124]
c_{max} [σ = 0.077]
c_{min} [σ = 0.087]
c_{max} [σ = 0.06]
c_{min} [σ = 0.067]
c_{max} [σ = 0.048]
c_{min} [σ = 0.069]
c_{max} [σ = 0.092]
c_{min} [σ = 0.115]
c_{max} [σ = 0.085]
c_{min} [σ = 0.138]
c_{max} [σ = 0.069]
c_{min} [σ = 0.09]
c_{max} [σ = 0.072]
c_{min} [σ = 0.368]
c_{max} [σ = 0.059]
c_{min} [σ = 0.07]
c_{max} [σ = 0.076]
c_{min} [σ = 0.103]
c_{max} [σ = 0.066]
c_{min} [σ = 0.086]
c_{max} [σ = 0.072]
c_{min} [σ = 0.088]
c_{max} [σ = 0.073]
c_{min} [σ = 0.099]
c_{max} [σ = 0.061]
c_{min} [σ = 0.106]



Supplementary Figure 7 | Confidence landscapes. The title of each plot indicates the experiment (E), the group (G) and the range of joint accuracies. The axes show the mean confidence of the worse, $c_{\min}$ (x-axis), and the better group member, $c_{\max}$ (y-axis) – with this division determined by comparing their fitted levels of sensory noise. Each mean indexes a maximum entropy distribution associated with a specific mean confidence. The grey dot indexes the joint accuracy expected under the distributions associated with the group members' observed mean confidence, a_{maxent}. The white dot indexes the joint accuracy expected under confidence matching, a_{match}, here the pair of confidence distributions associated with the average of the group members' mean confidence. The black dot indexes the joint accuracy expected under the optimal solution, a_{opt}. The values of each landscape were normalised to the range 0 (blue) to 1 (red).



Supplementary Figure 8 | Confidence matching in Experiment 4. The plot shows the probability density function over our measure of confidence matching: $\Delta m = |c^{\text{isolated}}_{\text{participant}} - c_{\text{partner}}| - |c^{\text{social}}_{\text{participant}} - c_{\text{partner}}|$. The empirical observations (x-axis) are overlaid, with each dot corresponding to a group. The difference in mean confidence was smaller in the social blocks than prior to interaction ($|c^{\text{social}}_{\text{participant}} - c_{\text{partner}}| < |c^{\text{isolated}}_{\text{participant}} - c_{\text{partner}}|$: $t(151) = -5.066$, $p < .001$, paired).



Supplementary Figure 9 | Serial dependence in confidence. a, Mutual influence. The axes show the influence of the partner's confidence on trial $t-1$ on the participant's confidence on trial t. Each dot is a group. The line is the best-fitting line of a robust regression; because the sorting of group members into 1 and 2 is arbitrary, we show the *p*-value for the slope of the best-fitting line averaged across 10^5 separate regressions, for each randomly re-labelling the members of a group as 1 and 2. **b**, Short-range serial dependence in isolated task. The y-axis shows coefficients from a linear regression encoding the degree to which a participant's confidence on trial t depended on their partner's confidence on trial $t-5$ to $t-1$ in the social task. We tested significance by comparing the coefficients pooled across participants to zero (trial $t-3$ to $t-1$: all $t(103) < 1.6$, all $p > .010$, one-sample *t*-test, null: 0). We included the stimulus on trial $t-5$ to t and the participant's own confidence on trial $t-5$ to $t-1$ as nuisance predictors. The data is from EXP1-I, EXP5-I and EXP6-I. **c**, Noisy read-out. We used our learning model (see **Methods**) to simulate data under different degrees of serial dependence (i.e., different adaptation rates, x-axis) and applied a standard measure of metacognitive ability to the data, meta-*d*' (see REFs in main text). This measure identifies the level of sensory noise that corrupts 'first-order' decision performance, *d*', and then asks how much more noise is needed to account for 'second-order' confidence performance, meta-*d*'. The standard interpretation is that such additional noise reflects a noisy read-out of the first-order information and thus lower metacognitive ability. The plot shows that higher adaptation rates are associated with higher differences between *d*' and meta-*d*' (y-axis). Importantly, by design, there is no noisy read-out in our learning model; only the mapping function is changing. It would therefore be wrong to attribute a higher difference between *d*' and meta-*d*' to a noisy read-out; the read-out is only noisy from the perspective of the experimenter. In a way, differences between *d*' and meta-*d*' can reflect higher metacognitive ability when they are driven by dynamic changes to the mapping function that make sense given the current context. In our simulations, we assumed that that a pair of agents had the same levels of sensory noise ($\sigma = .10$); that their mapping functions were updated so as to maintain maximum entropy over confidence; and that the learning rate was fixed ($\alpha = .12$) for both agents. In each simulated experiment, the agents performed 160 trials, with stimuli drawn as in our task. To control for random response variation due to sensory noise, we averaged across 10^3 simulated experiments.

data		condition	*mean* accuracy (*SD*)	*mean* confidence (*SD*)
GROUP 1	**participant**	*isolated*	.711 (.052)	2.697 (.613)
		ALCL	.727 (.062)	3.079 (.649)
		ALCH	.717 (.068)	3.566 (.675)
		AHCL	.714 (.065)	2.659 (.538)
		AHCH	.720 (.053)	3.313 (.610)
	virtual partner	*ALCL*	.653 (.054) *$	2.427 (.092) *$
		ALCH	.659 (.053) *$	4.582 (.096) *$
		AHCL	.792 (.059) *$	2.417 (.069) *$
		AHCH	.784 (.054) *$	4.572 (.097) *$
GROUP 2	**participant**	*isolated*	.715 (.042)	2.832 (.702)
		ALCL	.713 (.056)	3.121 (.776)
		ALCH	.730 (.055)	3.685 (.669)
		AHCL	.726 (.045)	2.796 (.512)
		AHCH	.719 (.063)	3.412 (.646)
	virtual partner	*ALCL*	.639 (.046) *$	2.043 (.081) *$
		ALCH	.666 (.051) *$	4.083 (.100) *$
		AHCL	.800 (.033) *$	2.046 (.056) *$
		AHCH	.789 (.041) *$	4.023 (.067) *$

Supplementary Table 1 | Manipulation checks for Experiment 4. The summary statistics are shown separately for the first and the second group of participants (19 in each group). We used two generic confidence distributions to specify the confidence of the computer-generated partners. For the first group of participants, we used the following confidence distributions: $p_{\mathrm{low}} = [.35,\ .27,\ .15,\ .11,\ .08,\ .04]$ with *mean* = 2.42 and $p_{\mathrm{high}} = [.04,\ .08,\ .11,\ .15,\ .27,\ .35]$ with *mean* = 4.56 – where the first element corresponds to confidence level 1, the second element corresponds to confidence level 2, and so forth. For the second group of participants, we used the following confidence distributions: $p_{\mathrm{low}} = [.44,\ .30,\ .14,\ .06,\ .02,\ .04]$ with *mean* = 2.04 and $p_{\mathrm{high}} = [.10,\ .12,\ .14,\ .16,\ .18,\ .26]$ with *mean* = 3.86. The second set of confidence distributions were set so as to better fit those of the first group of participants. We used paired *t*-tests (one-tailed) to test whether the data of participants were significantly different from that of the virtual partners. *: data of virtual partners significantly different from that of participants at baseline (isolated task). $: data of virtual partners significantly different from that of participants in a given social condition (social task).

	condition	gender (1=female)	like (0-100)	cooperation (0-100)	more accurate (1=yes)	more confident (1=yes)
GROUP 1	*ALCL*	0.75	65.25	70.40	0.20	0.13
	ALCH	0.26	57.15	69.00	0.25	0.73
	AHCL	0.55	72.50	75.10	0.73	0.15
	AHCH	0.20	58.00	69.85	0.65	0.90
GROUP 2	*ALCL*	0.65	73.17	67.50	0.33	0.06
	ALCH	0.24	61.11	64.83	0.31	0.81
	AHCL	0.76	77.89	81.39	0.75	0.14
	AHCH	0.24	70.00	76.11	0.72	0.86
COMBINED	*ALCL*	0.70	69.00	69.03	0.26	0.09
	ALCH	0.25	59.03	67.03	0.28	0.76
	AHCL	0.65	75.05	78.08	0.74	0.14
	AHCH	0.22	63.68	72.82	0.68	0.88

Supplementary Table 2 | Questionnaire data from Experiment 4. Participants were asked to complete a questionnaire about each partner after each social block. They were asked to indicate: (1) whether they thought the partner was male [0] or female [1]; (2) how much they liked the partner [0-100]; (3) how well they performed as a group [0-100]; (4) whether the partner was less [0] or more [1] accurate than they were; and (5) whether the partner was less [0] or more [1] confident than they were. The table shows the average responses. See Supplementary Table 1 for difference between the first and the second group of participants.