1 Supplementary Methods

1.1 Fitting procedures

Our Bayesian model fitting and comparison approach has been described in detail previously [58,60], but we repeat it here very briefly for completeness.

Each model contains, for each participant, a vector of parameters \( h \). The maximum a posteriori estimate of each parameter \( h_i = \text{argmax}_h p(A_i|h) p(h|\theta) \) is inferred for each participant \( i \) based on that participants’ actions \( A_i \). The parameters of the prior distribution \( \theta \) are inferred via approximate Expectation-Maximisation [59] and set to the maximum likelihood given all the data by \( all \) the \( N \) experimental sessions:

\[
\hat{\theta}^{ML} = \text{argmax}_\theta p(A|\theta)
\]

Here we let \( A = \{A_i\}_{i=1}^N \) be all the actions in all the experimental session. The approximation concerns the use of a Laplacian approximation for the E-step, letting:

\[
p(h|A_i) \approx \mathcal{N}(h_i^{(k)}, \Sigma_i^{(k)})
\]

\[
h_i^{(k)} = \text{argmax}_h p(A_i|h) p(h|\theta^{(k-1)})
\]

\( \mathcal{N}(\cdot) \) denotes a Gaussian distribution with variance \( \Sigma_i^{(k)} \) and mean \( h_i^{(k)} \). In the M-step, the group level mean parameters \( m \) and (factorized) variance \( v^2 \) are inferred simply as:

\[
m^{(k)} = \frac{1}{N} \sum_i h_i^{(k)}
\]

\[
(v^{(k)})^2 = \frac{1}{N} \sum_i \left[ (h_i^{(k)})^2 + \Sigma_i^{(k)} \right] - (m^{(k)})^2
\]

1.2 Model comparison

Models are compared based on the likelihood of the data given the model \( \log p(A|M) \) because we have no \textit{a priori} preference for any particular model. As previously, we approximate this quantity in two steps. At the group level [43], we utilize the standard Bayesian Information Criterion (BIC) approach:

\[
\log p(A|M) = \int d\theta p(A|\theta) p(\theta|M) \\
\approx -\frac{1}{2} \text{BIC} = \log p(A|\hat{\theta}^{ML}) - \frac{1}{2} |M| \log(|A|)
\]

Here, \( |A| \) is the total number of choices summed over all experimental sessions while \( |M| \) is the number of prior parameters fitted (note that there are one group-level mean and one group-level variance for each parameter).
Notably, though, \( \log p(A|\hat{\theta}^{ML}) \) is the sum of integrals over the individual parameters (hence the prescript “\( i \)” to BIC). We evaluate this by sampling from the empirical prior and averaging over the samples:

\[
\log p(A|\hat{\theta}^{ML}) \approx \sum_{i} \log \frac{1}{K} \sum_{k=1}^{K} p(A_{i} | h^{k})
\]

\[
h^{k} \sim p(h|\theta^{ML})
\]

1.3 Belief model specification

The model 'Belief' applies the belief variable \( \zeta \) only to the learned \( Q \) values, and not to the instructions \( I \). Prima facie this is surprising, as if a subject is able to correctly identify the stimulus to follow instructions, then they should be able to do so for the learned values, too. Here we briefly show that the models are, in fact, equivalent.

In model 'Belief', choices are determined by the term in the softmax:

\[
W_{t}(a_{t}, s_{t}) - W_{t}(\bar{a}_{t}, s_{t}) = \gamma(I(a_{t}, s_{t}) - I(\bar{a}_{t}, s_{t})) + f(Q)
\]

where \( f(Q) \) is a term that only involves \( Q \) and so is not important for the effect of \( \zeta \) on \( I \).

An alternative model, where \( \zeta \) applies to both the learned component \( Q \) and the instructions \( I \), would read:

\[
W_{t}(a_{t}, s_{t}) = \gamma(\zeta I(a_{t}, s_{t}) + (1 - \zeta) I(\bar{a}_{t}, s_{t})) + \zeta Q_{t}(a_{t}, s_{t}) + (1 - \zeta) Q_{t}(a_{t}, \bar{s}_{t})
\]

and so the term in the choice-determining softmax would be

\[
W_{t}(a_{t}, s_{t}) - W_{t}(\bar{a}_{t}, s_{t}) = \gamma(\zeta I(a_{t}, s_{t}) + (1 - \zeta) I(\bar{a}_{t}, s_{t})) - \gamma(\zeta I(\bar{a}_{t}, s_{t}) + (1 - \zeta) I(\bar{a}_{t}, s_{t})) + f(Q)
\]

however, \( I(a_{t}, s_{t}) = I(\bar{a}_{t}, \bar{s}_{t}) \), and similarly for the off-diagonal terms of \( I \), so we have

\[
W_{t}(a_{t}, s_{t}) - W_{t}(\bar{a}_{t}, s_{t}) = \gamma((2\zeta - 1) I(a_{t}, s_{t}) - (2\zeta - 1) I(\bar{a}_{t}, s_{t})) f(Q)
\]

which is this equivalent to redefining \( \gamma \) as \( \gamma(2\zeta - 1) \).

Another alternative, would be to argue that uncertainty affects not only the selection of the appropriate \( Q \) value, but also the learning. To test this legitimate idea, we fitted an additional model in which, on each
iteration, the $Q$ value of the chosen action $a_t$ for both the presented stimulus $s_t$ and the other stimulus $\bar{s}_t$ is updated to an extent proportional to the belief:

$$Q(a_t, s_t) = Q(a_t, s_t) + \zeta \epsilon (\rho r_t - Q(a_t, s_t))$$

$$Q(a_t, \bar{s}_t) = Q(a_t, \bar{s}_t) + (1 - \zeta) \epsilon (\rho r_t - Q(a_t, \bar{s}_t))$$

This model performed very similarly to the model 'Action' and performed worse than the model 'Belief' (log Bayes factor compared to model 'Belief' 138.04). We therefore did not analyse it further.
2 Supplementary Results

2.1 Model comparison for Stress dataset

Figure S1: Model comparison log Bayes factors. Left panel is the same as in Figure 2A for comparison; the middle panel for all but the Stress dataset; and the right panel for the Stress dataset alone. Throughout, there is clear evidence for model ‘Belief’.

Because they faced a more difficult perceptual task (a smaller difference between the ‘long’ and the ‘short’ mouth), participants in the stress dataset, showed on average a lower probability of correct choice, and were consequently less well predicted by all models (c.f. Figure 2). We therefore performed additional, separate model fits and comparisons for the Stress dataset, and the remaining five datasets; and for all datasets separately. Figure S1 shows the results for the Stress dataset and all other datasets combined. The same ‘Belief’ model provided the (jointly) best fit in all datasets. In addition, the belief values did not differ between the Stress dataset and the other datasets (data not shown). Combined with the finding that participants who responded correctly less often had lower $\gamma$ values (Figure 2C), this shows that these participants specifically assigned less weight to the instructions, but not that they differed in their ability to identify, or learn about, the stimuli. Thus, the accuracy difference between the Stress dataset and the other datasets was well captured by the parameter $\gamma$ of the model and does not require separate treatment of the Stress dataset.
2.2 Surrogate data comparison

Figure S2: Evolution of choice bias for all datasets (left to right) and all models (top to bottom). Each dataset is split into two panels, showing the ‘control’ and ‘treatment’ subgroups in green and red, respectively. Each curve shows the difference in the log probability of a correct response to the rich minus the lean stimulus, averaged across participants. The black curves show the same measure averaged across the 70 surrogate datasets. Note that these models were fit to all six datasets jointly, yet all datasets are fitted reasonably well.

As additional test of model fit, we generated data from the model. Sampling is a good guide as to whether the model captures the data; but does not measure parsimony. We generated 70 surrogate datasets from each models for direct comparison with the experimental results on the evolution of the choice bias log p(correct response—rich stimulus) - log p(correct — lean stimulus). We obtained the maximum a posteriori (MAP) parameters $h^k$ for each participant, and then used this together with the original sequence of stimuli for that participant to generate a novel sequence of 300 choices, although we used a truly random reward sequence rather than the pseudorandom sequence used in the experiment to generate this data. Because datasets were pooled for parameter inference, we expect the resampled data to regress towards the mean of all datasets (this effect should be stronger in smaller datasets). However, we should expect roughly to replicate the general pattern observed in the raw data.

Figure S2 shows the evolving bias towards the rich stimulus for all models and all datasets. The evolution of the bias was reasonably well captured across all datasets. It is notable that the model 'Action' seems to perform similarly to the model 'Belief' despite being simpler. We therefore repeated the regression analyses with the parameters derived from this model.