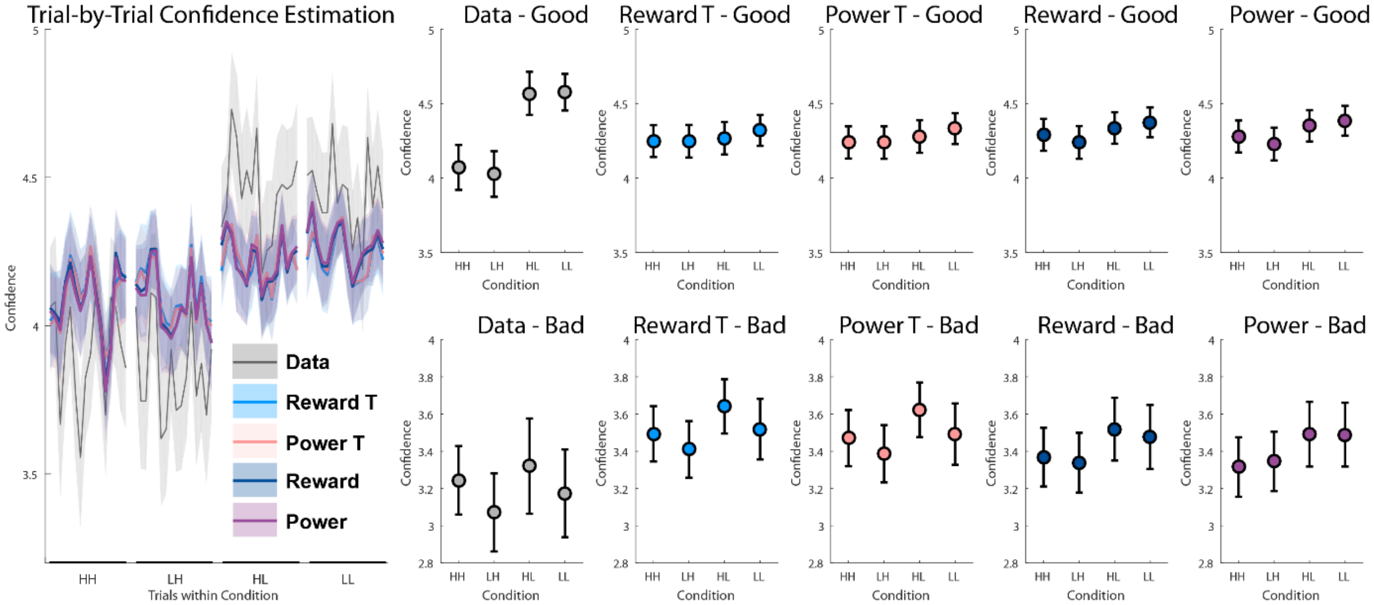

S7 Fig. Power Utility Model Performance in Experiment 1

To examine other utility functions we fitted a utility model which transforms the rewards in each trial according to the power utility function [1,2]:

$$\mathrm{U(R)} = \frac{R^{(1-\gamma)}}{1-\gamma} \quad 0 \le \gamma < 1$$

Where γ is the risk aversion factor – the closer it is to 1 the participant is more risk averse (and the closer the function is to log(r)). We used a model that learns from these transformed values, i.e. from utilities and not directly from the rewards, but was otherwise exactly the same as the ‘Reward’ model:

$$\begin{cases} Q_a(t+1) = Q_a(t) + \alpha\big(U\big(R(t)\big) - Q_a(t)\big) \\ \qquad\qquad Q_b(\mathrm{t}+1) = \mathrm{Q}_b(t) \end{cases}$$

When option a is chosen, its value is updated according to the difference between its current value and the utility of the option’s current reward, with a learning rate α. Decision in each trial was then carried using a softmax rule.

We fitted this ‘Power’ model to the choice data, and an additional ‘Power-T’ model which added the drift to threshold of the unchosen option mechanism:

$$\begin{cases} Q_a(t+1) = Q_a(t) + \alpha\big(U\big(R(t)\big) - Q_a(t)\big) \\ \quad Q_b(\mathrm{t}+1) = \mathrm{Q}_b(t) + \alpha\big(T - Q_b(t)\big) \end{cases}$$

We examined the fit of these models to the data and how well they predicted the confidence reports.

We found that the ‘Power’ and ‘Power-T’ models performed very similarly to the ‘Reward’ and ‘Reward-T’ models respectively. Their WAIC values were: ‘Power’ 228.06 ± 67.46, ‘Power-T’: 214.51 ± 68.66, whereas the ‘Reward’ models WAIC values were: ‘Reward’ 226.55 ± 67.67, ‘Reward-T’ 214.57 ± 68.79. ‘Power-T’ was as good as our best model in explaining choice behaviour.

We than examined how well the ‘Power’ models explained confidence reports. Again, they fared similarly to the ‘Reward’ models with linear fit (R^2) of: Power’ 0.21 ± 0.22, ‘Power-T’: 0.21 ± 0.21, whereas the ‘Reward’ models WAIC values were: ‘Reward’ 0.21 ± 0.22, ‘Reward-T’ 0.21 ± 0.21.

Finally, we examined the predicted confidence reports in the four condition blocks, and found that the patterns predicted by the ‘Power-T’ model were identical to the pattern predicted by the ‘Reward’ model. We concluded that the transformation of reward in a trial by trial manner did not introduce any new mechanism to learning beyond the one already implemented by the ‘Reward’ model.