**Supplementary File 4. Causal inference model inference and fitting details**

In order to infer the posterior over Dimp in the implicit task (Eq. 14), we can combine Eqs. 5,6 and 11 to get the likelihood over Dimp and C by integrating over all inferred cue location values as follows,

$p\_{implicit}\left(X\_{a},X\_{v}|D\_{imp},C\right)∝∬\_{}^{}p\left(S\_{a}\right)p\left(S\_{v}\right)p\_{implicit}\left(S\_{a},S\_{v}|D\_{imp},C\right)dS\_{a}dS\_{v} $ **(Eq. S1)**

We can expand Eq. S1 using rules for the product of Gaussian probability density functions to get,

$$p\_{implicit}\left(X\_{a},X\_{v}|D\_{imp},C\right)∝\left(C-1\right)Φ\left[^{D\_{imp}\left(X\_{a}α\_{a}+μ\_{a}\left(1-α\_{a}\right)\right)}/\_{σ\_{a}\sqrt{α\_{a}}}\right]$$

$+\left(2-C\right)N\left(X\_{a'};X\_{v'},σ\_{a}^{2}α\_{a}+σ\_{v}^{2}α\_{v}\right)Φ\left[^{D\_{imp}\left(X\_{a'}α\_{a^{'}v^{'}} +X\_{v'}(1-α\_{a^{'}v^{'}})\right)}/\_{σ\_{a}\sqrt{α\_{a^{'}v^{'}}α\_{a}}}\right] $ **(Eq. S2)**

where $α\_{a}=\frac{σ\_{ap}^{2}}{σ\_{a}^{2}+σ\_{ap}^{2}}$, $α\_{v}=\frac{σ\_{vp}^{2}}{σ\_{v}^{2}+σ\_{vp}^{2}}$, $X\_{a'}=X\_{a}α\_{a}+μ\_{a}\left(1-α\_{a}\right)$, $X\_{v'}=X\_{v}α\_{v}+μ\_{v}\left(1-α\_{v}\right)$ and $α\_{a'v'}=\frac{σ\_{v}^{2}α\_{v}}{σ\_{a}^{2}α\_{a}+σ\_{v}^{2}α\_{v}}$

which we can evaluate using the previous equations as

$p\_{implicit}\left(X\_{a},X\_{v},C=c\right)∝p\_{implicit}(X\_{a},X\_{v}|D\_{imp}=1,C=c)p(C=c|D=1)p\_{implicit}(D\_{imp}=1)$ **Eq. (S3)**

and

$p(C=c|X\_{a},X\_{v})∝\sum\_{d\in \left\{-1,1\right\}}^{}p\_{implicit}\left(D\_{imp}=d,C=c\right)p(C=c|D\_{imp}=d)p\_{implicit}(D\_{imp}=d)$ **(Eq. S4)**

The subject uses the inferred posterior probability over Dimp to make their choice by comparing which response has a higher probability of being correct (Eq. 14). Since Dimp is a binary variable, this is equivalent to checking if $p\left(X\_{a},X\_{v}\right)>0.5$. Since we do not have access to the subject’s sensory observations, we marginalize over all possible sensory observations using Eqs. 3 and 4. We also model lapses in subject judgements using a lapse rate (denoted by $p\_{lapse rate}^{imp}$) and lapse bias (denoted by $p\_{lapse bias}^{imp}$). Lapse rate characterizes the probability that the subject makes a choice independent of the stimulus. The lapse response is modeled as a random sample from a Bernoulli distribution with parameter $p\_{lapse rate}^{imp}.$ Mathematically, the probability of the subject making a choice $R\_{implicit}$=1 (equivalently for $R\_{explicit}$=1) for a given set of experimenter defined $ϵ\_{a} $ and $ϵ\_{v}$ is given by,

$p\left(ϵ\_{a},ϵ\_{v}\right)=p\_{lapse rate}^{imp}p\_{lapse bias}^{imp}+\left(1-p\_{lapse rate}^{imp}\right)∬\_{}^{}H[p\left(X\_{a},X\_{v}\right)-0.5]p\left(ϵ\_{a}\right)p\left(ϵ\_{v}\right)dX\_{a}dX\_{v} $

**(Eq. S5)**

We perform the integral in Eq. S5 for the implicit task by numerically evaluating the decision boundary using a multi-dimensional bisection method (Bachrathy et al. 2012) and then using this to evaluate the integral analytically (the complete analytic solution is unavailable). This provides a better estimate than traditional sampling or numerical methods since the decision function is not smooth. To verify the estimates, we also obtained estimates by numerical integration using gaussian quadrature (Golub et al. 1969). We use analytic solutions for the explicit task.

For the aggregate subject model predictions, we found the Maximum a Posteriori (MAP) estimate for the model parameters under weakly informative priors for the sensory parameters and choice parameters (details in the *Supplementary Material S6*). The fitting was done using a quasi-newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) unconstrained optimization procedure (fminunc in MATLAB) using 40 restarts to find the global minimum. We obtained full the posterior over the parameters for the aggregate control subjects using generalized elliptical slice sampling (Nishihara et al. 2014) which allowed us to quantify the uncertainty over the parameter estimates. Since our model prediction for ASD subjects requires setting the sensory parameters to the same value as the control and explaining the differences in terms of the choice and inference parameters, we obtained posterior estimates for the choice and inference parameters for $n\_{approx}$ independent samples from the control parameter fits (184 in our case) and then computed the effective mean and standard deviation across each posterior to get the uncertainty around the fitted ASD parameters. The posterior over the ASD parameters obtained using the method described above is mathematically illustrated below

$$p\left(D\_{control},D\_{ASD}\right)=…∭\_{}^{}p\left(D\_{control},D\_{ASD},θ\_{control}^{''}\right)p\left(D\_{control}\right)dθ\_{control}^{''}$$

$p\left(D\_{control},D\_{ASD}\right)≈\frac{1}{n\_{approx}}\sum\_{θ\_{control}^{''\left(i\right)}∼p\left(D\_{control}\right)}^{}p\left(D\_{control},D\_{ASD},θ\_{control}^{''\left(i\right)}\right) $**(Eq. S6)**

where $θ\_{ASD}^{'}$ are the choice and inference parameters for the aggregate ASD subject, $θ\_{control}^{''}$ are the sensory parameters for the aggregate control subject, $D\_{ASD},D\_{control}$ are ASD and control data.

In order get a goodness of fit estimate of the model, we computed the explainable variance explained (EVE, Haefner et al. 2008). This estimate is an extension of the tradition variance explained/ coefficient of determination but accounts for the uncertainty in the data generation process which is the case for us with limited number of trials per condition. In addition, it corrects for the number of parameters in the model to allow for overfitting.

We also obtained full posteriors over model parameters for the individual subjects by jointly fitting the model to all experiments with weakly informative priors. We modeled each observer population with a hierarchical model where an individual observer’s parameters are independent draws from the population parameter, parameterized as a Gaussian, i.e. $θ\_{subject}^{ }∼N(θ\_{population},σ\_{population}^{2})$ where $θ$ are the parameters of the causal inference model. We also approximated the posterior over the subject parameters as Gaussians which allowed us to analytically combine the individual posteriors to a combined Gaussian population posterior for further hypothesis testing and obtaining confidence intervals (CI). For the individual subject estimates, we plot 68%CI. We obtained the p-value for the difference between ASD and control subjects parameters using two methods: (a) Normal approximated analytical p-value which assumes that the sample variance is equal to the population variance, (b) Welch t-test which relaxes the above assumption. Both methods gave comparable p-values and we conservatively considered the higher p-value for significance testing.