Supplementary Materials for

A frontal cortical world model shapes inference of causal relationships in hippocampus

Shuyi Luo[1*], Muchen Li[1], Deng Pan[1], Naomi Kingston[1], Miriam C. Klein-Flügge[1,2,3], Ali Mahmoodi[1†], Matthew F.S. Rushworth[1,2 *†]

Corresponding authors: shuyi.luo@psy.ox.ac.uk, matthew.rushworth@psy.ox.ac.uk

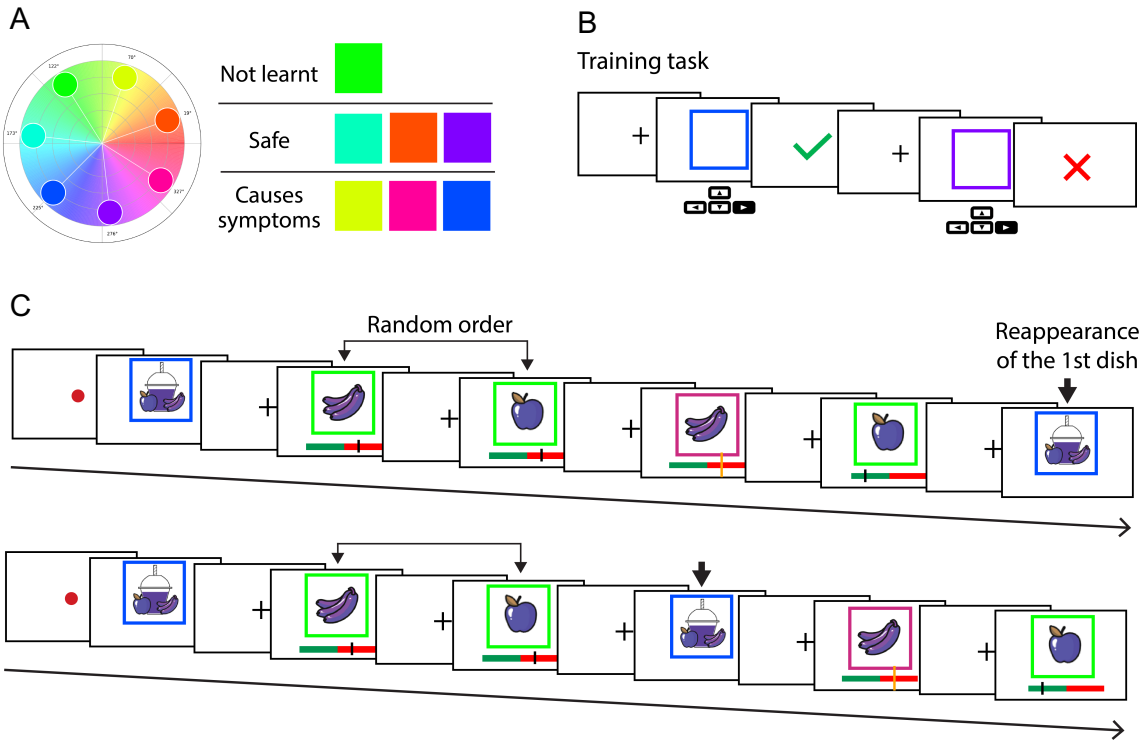

Fig. S1. Supplementary information on task design, related to Fig. 1B. (A) To minimise colour-related bias and facilitate representational similarity analysis (RSA) of neural data, seven colours with equal hue spacing were selected. Colours representing the two causal outcomes ("causes symptoms" and "safe") were matched in overall hue distance from a neutral colour indicating "not learnt." **(B)** Participants completed an online training task one day prior to the fMRI session, followed by a refresher training session on the day of scanning, to learn the meanings of the colour frames (see Methods for details). **(C)** After presentation of the first dish, participants rated both the experienced and inferred objects in a randomised order. The first dish either reappeared at the end of the trial (top) or before the presentation of the experienced object (bottom). This design balanced the temporal distance between the first dish and both the experienced and inferred objects across trials.

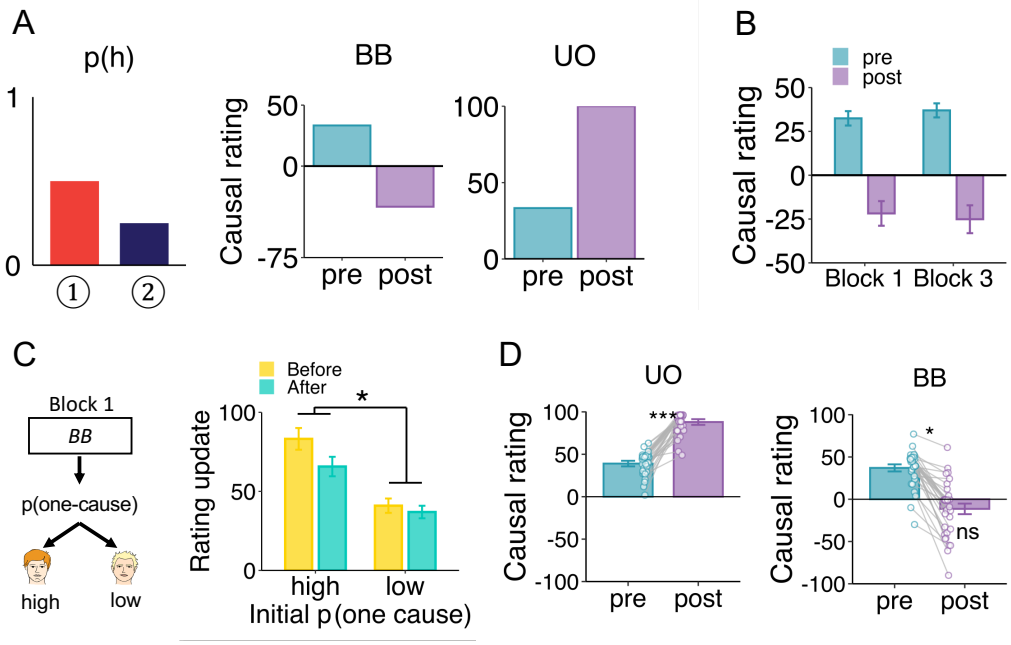

Fig. S2. Supplementary behavioural results, related to Fig. 1D-G. (A) Simulation results from the Bayesian model showing rating updates of the inferred object in both BB and UO trials under a relatively strong prior for a one-cause structure, closely resembling the observed participant data. **(B)** BB effect before and after UO trials (block 2), showing a marginal increase in the causal rating of the inferred object following UO trials (B=-7.58, SE=4.45, p=.089). **(C)** Participants were divided based on their post-ratings of the inferred object in the first BB trial block into groups with high versus low initial beliefs in one-cause hypothesis. The right panel shows rating updates in BB trials before and after the two-cause trials for each group, indicating a larger change in participants with a stronger initial belief in a one-cause hypothesis (B=13.45, SE=5.38, p=.018). **(D)** The experiment comprised two sessions. Prior to the second session, participants had already been exposed to two-cause trials in the first session, which appeared to have a persistent effect on their inference in BB trials. Specifically, participants no longer made strong inferences about the inferred object in BB trials (right), even before re-exposure to two-cause trials, in contrast to their initial performance (Fig. 1E). This suggests that exposure to two-cause trials appeared to shift participants' beliefs toward a two-cause hypothesis in a relatively persistent manner. In contrast, participants continued to show a robust UO effect (left). In panels B-D, the error bars represent standard error of the mean (SEM), and each dot represents an individual participant. ***p<.001, **p<.01, *p<.05, ns not significant.

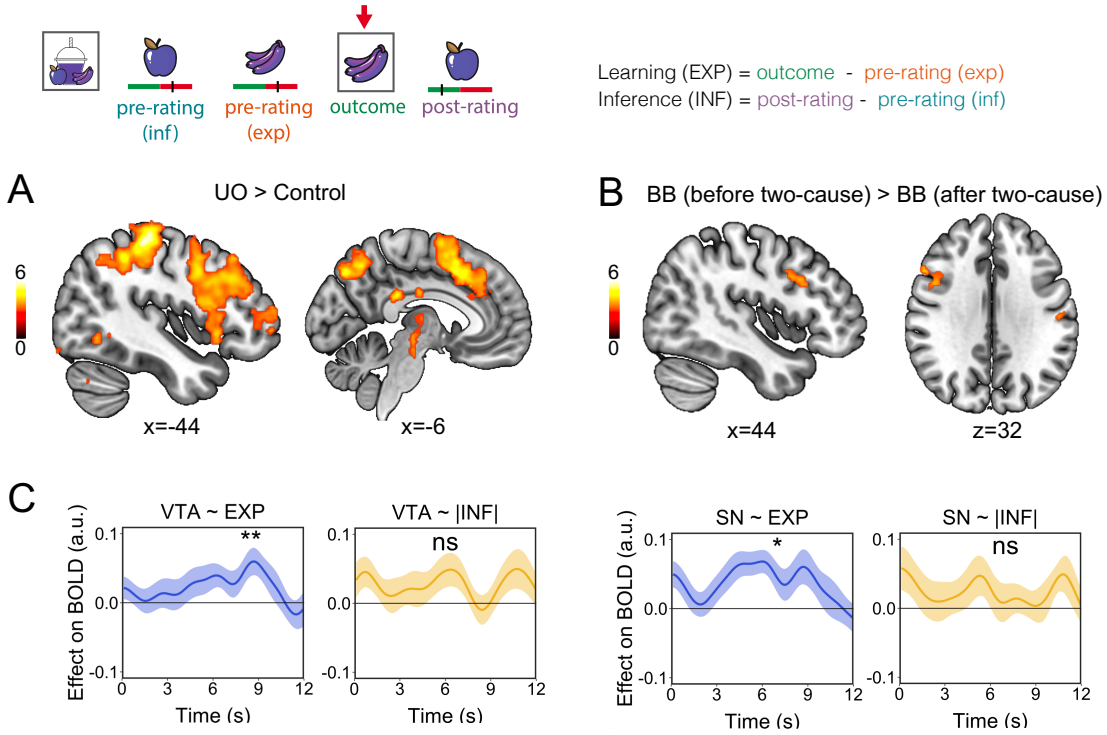

Fig. S3. Supplementary univariate results, related to Fig. 2B-C. (A) Whole-brain GLM comparing brain activity during the revaluation phase (i.e., when the outcome of the experienced object was presented) between UO and control trials. Behaviourally, both learning and inference were stronger in UO than in control trials; thus, these results likely reflect regions involved in both processes. **(B)** Whole-brain GLM comparing brain activity during the revaluation phase in BB trials before and after two-cause trials. Because the causal evidence was identical, learning-related processing remained constant, whereas inference was altered following two-cause trials, as reflected in the reduced rating update of the inferred object (Fig. 1G). These results further support the role of the inferior frontal sulcus (IFS) in inference. **(C)** Time-course analysis using independently defined masks of the ventral tegmental area (VTA) and substantia nigra (SN), showing that activity in these regions was associated with the experienced update but not the inferred update.
All whole-brain results shown were obtained after whole-brain cluster correction with a voxel-wise threshold of $Z>3.1$ and a cluster significance threshold of $p<.05$. In C the lines and shaded areas show the mean and standard error of the mean (s.e.m), respectively. $^{**}p<.01$, $^{*}p<.05$, ns: not significant.

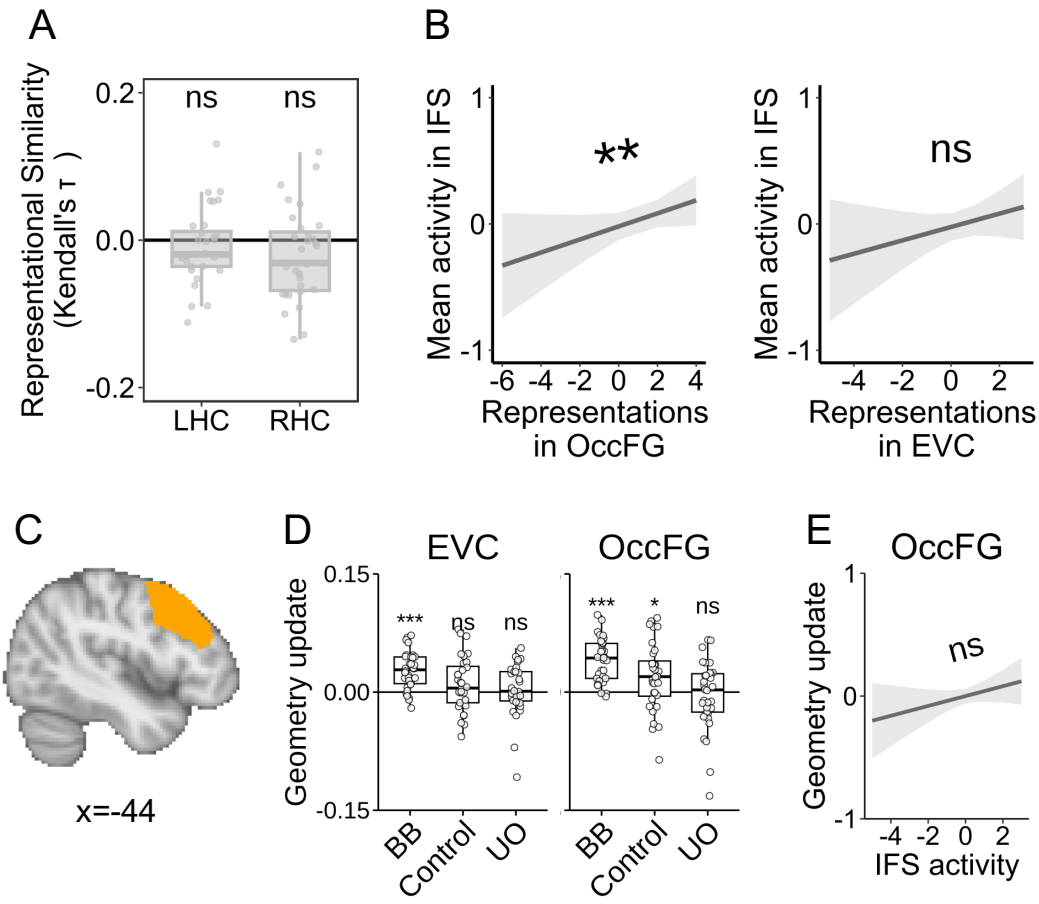

Fig. S4. Supplementary multivariate analysis, related Fig. 2E-J. (A) The representation of the object identity was not found in the HC. **(B)** The relationship between representations of the inferred object in the occipital fusiform gyrus (OccFG) and IFS activity (as shown in Fig. 2D) was replicated using an alternative similarity metric (Mahalanobis distance; B=0.06, SE=0.02, p=.002). No such relationship was observed between early visual cortex (EVC) representations and IFS activity (B=0.05, SE=0.05, p=.271). **(C)** The middle frontal gyrus (MFG) mask from the Harvard–Oxford Cortical Structural Atlas (Desikan et al., 2006) was used as an independently defined anatomical mask to approximate the IFS, as it encompassed the IFS activation associated with the inference. **(D)** In contrast to the HC, neither EVC nor OccFG showed the predicted pattern of geometric change–an increase in the distance ratio for BB trials, a decrease for UO trials, and no change in control trials. **(E)** No association was observed between IFS activity and representational geometry update in OccFG (B=0.04, SE=0.03, p=.191), in contrast to the HC (Fig. 2J).

In all boxplots, the horizontal band inside the boxes indicates the median and the whiskers extend to the most extreme data samples that lie within the intervals spanning 1.5 times the inter-quartile range. Each dot represents an individual participant. In B and E, lines and shaded areas indicate model estimates and standard errors derived from fitted multilevel models. ***p<.001, **p<.01, *p<.05, ns: not significant.

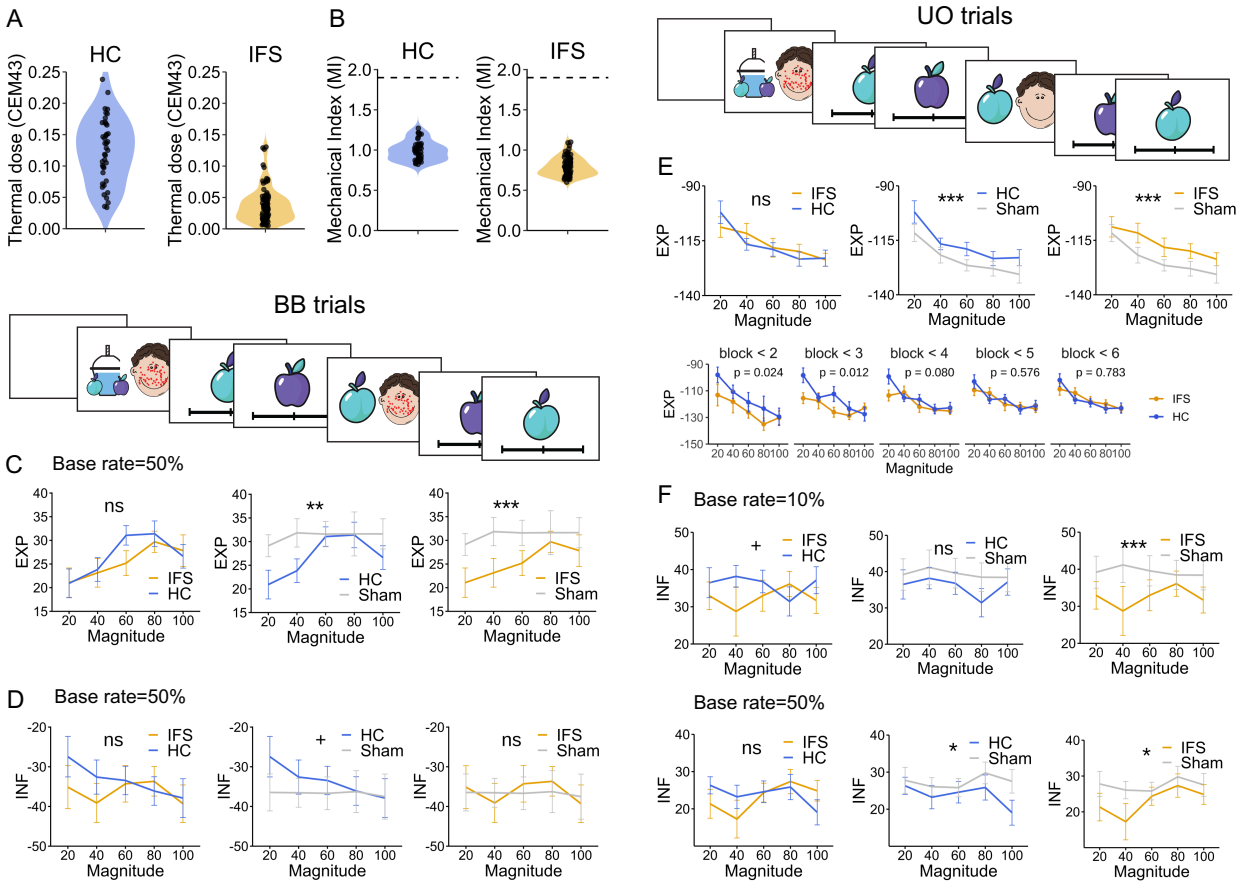

Fig. S5. Supplementary TUS results, related to Fig. 3. (A) Thermal simulations. Violin plots show the distribution of thermal dose (CEM43) across all HC and IFS sonications, with all values below 0.25. **(B)** Distribution of mechanical index (MI) across all sonications; all sonications had MI<1.9. **(C)** Supplementary to Fig. 3F, which showed a stronger decrease in experienced updates following HC stimulation under the low base-rate condition in BB trials (example BB trial illustrated above). Under the high base-rate condition, participants could not rely solely on observed outcomes to update their causal judgments of the experienced object. Instead, they needed to infer whether the observed allergic reaction reflected the causal property of the experienced object or the character's general propensity for allergic reactions regardless of what they consumed. Consequently, rating updates of the experienced object reflected a combination of learning and inference processes. The results indicate that both HC and IFS stimulation led to a reduction in such updates under this condition. **(D)** Supplementary to Fig. 4G–H. Whereas the main results showed decreased inferred updates following HC stimulation and altered inference patterns after IFS stimulation under the low base-rate condition, here we show that under the high base-rate condition, differences between stimulation conditions were attenuated. This may be because participants were already unable to reliably perform inference under the high base-rate condition even in the sham condition, such that further disruption of inference-related regions did not produce additional behavioural effects. **(E)** In UO trials (see illustration above), experienced updates were reduced in both HC and IFS conditions compared to sham, with no overall difference between HC and IFS. However, across blocks, experienced updates in the HC condition were significantly lower than in the IFS condition in the first half of the task (i.e., closer to the stimulation period), suggesting a stronger disruption of learning following HC stimulation. **(F)** In UO trials, inferred updates were significantly reduced in the IFS condition compared to sham, but not in the HC condition,

under the low base-rate condition. Under the high base-rate condition, both HC and IFS conditions showed reduced inferred updates, with no difference between them.
In A-B, the shaded area showed the distribution and each dot represents one sonication. In C-F, the lines and error bars show the mean and standard error of the mean (s.e.m), respectively. ***p<.001, **p<.01, *p<.05, ^{+}p<.10, ns: not significant.

Table S1. Clusters that significantly correlated with the experienced update during the revaluation phase, related to Fig. 2B.

			Peak MNI coordinates		
Cluster size	**P**	**Peak Z**	**x**	**y**	**z**
8785	0.0	5.0	-16	2	52
1514	2.13e-15	4.19	6	-64	52
892	1.28e-10	4.58	-36	-54	-30
643	2.1e-08	4.32	36	20	-6
530	2.38e-07	4.3	38	-54	-30
372	1.25e-05	3.93	-56	-40	30
232	0.0006	4.41	-12	-74	10
171	0.0044	3.77	46	-70	38
139	0.0135	3.82	24	-2	8

Table S2. Clusters that significantly correlated with the magnitude of the inferred update during the revaluation phase, related to Fig. 2C.

			Peak MNI coordinates		
Cluster size	**P**	**Peak Z**	**x**	**y**	**z**
277	1.86e-05	3.84	18	-84	-8
262	3.15e-05	4.06	-46	14	34
207	0.0002	3.90	36	-56	38
193	0.0004	3.83	38	10	26
136	0.0043	3.88	6	-74	46
133	0.0049	3.91	-8	16	50
111	0.0133	4.10	-36	0	60
105	0.0176	4.11	36	-62	-28
101	0.0213	3.90	-32	-62	42

Table S3. Clusters that showed significant repetition suppression effects, related to Fig. 2D.

			Peak MNI coordinates		
Cluster size	**P**	**Peak Z**	**x**	**y**	**z**
13152	0.0	6.6	-22	-82	-12
699	5.96e-08	4.42	-2	-58	32
411	1.69e-05	4.73	-22	-68	34
167	0.0097	4.47	-16	-34	-4

Table S4. In situ exposure parameter estimates from acoustic simulations in k-Plan, related to Fig. 3.

	Left HC					Right HC				
ID	**Free-field I_{sppa} (W/cm2)**	**Focal Depth (mm)**	**Target I_{sppa} (W/cm2)**	**Mechanical Index (MI)**	**CEM43**	**Free-field I_{sppa} (W/cm2)**	**Focal Depth (mm)**	**Target I_{sppa} (W/cm2)**	**Mechanical Index (MI)**	**CEM43**
1	75	107	5.07	0.97	0.1675	75	109	4.07	0.98	0.1099
2	75	110	4.99	0.89	0.0690	75	110	6.06	0.87	0.0715
3	75	110	5.66	0.86	0.0574	75	106	4.00	0.82	0.0489
4	75	117	5.44	0.84	0.0973	75	119	5.62	1.27	0.1904
5	75	108	9.27	0.86	0.0414	75	107	12.38	0.86	0.0356
6	75	113	4.65	1.06	0.1691	75	113	4.06	1.07	0.1466
7	75	107	4.65	0.92	0.1068	75	107	3.97	1.02	0.0760
8	75	117	5.40	0.93	0.0985	75	119	5.97	0.89	0.1833
9	75	113	6.23	1.20	0.1701	75	117	4.17	1.07	0.0338
10	75	114	4.53	0.89	0.1184	75	118	4.09	1.01	0.1914
11	75	120	4.86	1.08	0.1387	75	119	4.33	1.07	0.1503
12	75	117	3.73	1.14	0.1287	75	117	4.86	1.21	0.1653
13	75	109	5.03	0.83	0.1052	75	122	7.44	1.03	0.2381
14	75	112	4.14	0.96	0.0661	75	122	4.55	1.21	0.1765
15	75	122	7.21	1.03	0.1515	75	124	9.31	0.95	0.1361
16	75	117	6.19	1.05	0.0540	75	118	5.61	0.85	0.1438
17	75	108	5.10	1.01	0.1363	75	117	3.59	1.01	0.2173
18	75	119	6.40	0.99	0.1848	75	117	5.43	0.87	0.0894
19	75	119	4.41	1.20	0.1171	75	117	6.62	1.16	0.1009
20	75	115	5.41	0.96	0.1482	75	115	5.86	0.90	0.1251

		Left IFS					Right IFS				
ID	**sonication**	**Free-field I_{sppa} (W/cm2)**	**Focal Depth (mm)**	**Target I_{sppa} (W/cm2)**	**Mechanical Index (MI)**	**CEM43**	**Free-field I_{sppa} (W/cm2)**	**Focal Depth (mm)**	**Target I_{sppa} (W/cm2)**	**Mechanical Index (MI)**	**CEM43**
1	1	65	40.3	4.26	0.70	0.0091	65	44.0	3.56	0.65	0.0490
1	2	65	44.0	4.76	0.79	0.0467	65	45.9	5.46	0.74	0.0504
2	1	65	40.3	7.84	0.86	0.0281	65	42.2	6.00	0.80	0.0355
2	2	65	40.3	4.38	0.87	0.0338	65	42.2	6.59	0.79	0.0409
3	1	65	40.3	7.38	0.77	0.0150	65	42.2	8.34	0.75	0.0165
3	2	65	40.3	7.56	0.90	0.0475	65	42.2	9.47	0.76	0.0484
4	1	65	40.3	6.01	0.60	0.0083	65	44.0	7.46	0.83	0.0186
4	2	65	40.3	5.64	0.73	0.0497	65	44.0	7.38	0.67	0.0519
5	1	65	40.3	13.69	0.91	0.0078	65	40.3	8.15	0.70	0.0064
5	2	65	40.3	12.81	0.89	0.0166	65	40.3	7.35	0.79	0.0137
6	1	65	40.3	8.29	0.74	0.0051	65	40.3	7.55	0.67	0.0063
6	2	65	40.3	9.90	0.79	0.0112	65	42.2	8.30	0.71	0.0126
7	1	65	40.3	9.25	0.75	0.0173	65	44.0	7.13	0.71	0.0576
7	2	65	42.2	9.71	0.84	0.0553	65	45.9	7.17	0.66	0.0227
8	1	65	40.3	5.35	0.84	0.0189	65	45.9	9.29	0.75	0.0383
8	2	65	42.2	7.71	0.79	0.0362	65	45.9	8.77	0.78	0.0396
9	1	65	40.3	6.35	0.84	0.0140	65	44.0	6.44	0.68	0.0142
9	2	65	40.3	6.67	1.06	0.0468	65	44.0	4.44	0.65	0.0343
10	1	65	40.3	4.58	0.89	0.0200	65	40.3	5.26	0.91	0.0244
10	2	65	40.3	4.39	0.70	0.0229	65	43.0	5.17	0.75	0.0294
11	1	65	40.3	6.67	0.96	0.1018	65	44.9	7.06	0.68	0.1291
11	2	65	41.4	7.59	0.94	0.1268	65	45.9	6.55	0.66	0.1304
12	1	65	40.3	9.66	0.85	0.0376	65	44.9	8.32	0.73	0.0493
12	2	65	41.4	8.92	0.78	0.0474	65	45.9	7.25	0.69	0.0531
13	1	65	40.3	8.02	0.83	0.0189	65	45.4	8.79	0.73	0.0788
13	2	65	41.2	7.93	0.75	0.0764	65	45.7	8.70	0.72	0.0802
14	1	65	40.3	11.92	1.10	0.0220	65	42.2	12.81	0.88	0.0310
14	2	65	40.3	11.19	0.84	0.0294	65	44.0	14.31	0.93	0.0322
15	1	65	40.3	4.94	0.89	0.0134	65	42.6	6.78	0.67	0.0432
15	2	65	40.3	4.41	0.87	0.0412	65	44.4	6.82	0.69	0.0445
16	1	65	40.3	7.26	0.70	0.0057	65	40.9	5.22	0.64	0.0176
16	2	65	40.3	7.57	0.67	0.0159	65	42.2	6.81	0.64	0.0231
17	1	65	42.7	5.39	0.64	0.0113	65	45.9	7.41	0.70	0.0312
17	2	65	44.4	4.34	0.72	0.0287	65	48.7	8.01	0.69	0.0327
18	1	65	40.3	7.69	0.79	0.0110	65	42.0	8.65	0.81	0.0749
18	2	65	40.3	7.43	0.85	0.0728	65	42.3	7.78	0.78	0.0762
19	1	65	40.3	4.26	0.70	0.0091	65	44.0	3.56	0.65	0.0490
19	2	65	44.0	4.76	0.79	0.0467	65	45.9	5.46	0.74	0.0504
20	1	65	40.3	7.84	0.86	0.0281	65	42.2	6.00	0.80	0.0355
20	2	65	40.3	4.38	0.87	0.0338	65	42.2	6.59	0.79	0.0409