

Supplementary Methods



Figure S1. Mooney image examples

Note. The black/white abstracted images on the lower panel represent the Mooney images and the corresponding real-world pictures are presented on the upper panel.

Categories as response options for solution in the MRI scanner

1. Leisure & Hobbies,
2. Tool, weapon or technology,
3. Clothing & Accessories,
4. Vehicles & Road Traffic,
5. Landscape or building,
6. Vehicles & Road Traffic,
7. Predators or rodents,
8. Mammal,
9. Reptiles or Insects,
10. Birds or animal that lives in/near the water,
11. Events or Science,
12. Objects in the house,
13. Human Being

Relationship between different dimensions of insight experience

We explored whether it is justified to combine the three insight dimensions

(*Suddenness*, *Emotion* and *Certainty*) into one single measure. For this, we calculated a measurement model for a latent insight factor from those three insight dimensions for correctly identified Mooney images. The latent factor was estimated within a confirmatory factor analysis (CFA) in R using the lavaan package (v. 0.6-18)¹. The model fit was evaluated via Bentler's comparative fit index (CFI), root mean square error of approximation (RMSEA) and standardized root-mean-square residual (SRMR). Accepted thresholds indicating good model fit are $RMSEA \leq .05$, $SRMR < 0.1$ and $CFI \geq .95$.²⁻⁴ Note, due to a lack of degrees of freedom no exact χ^2 goodness-of-fit statistic could be calculated. Error variance of the latent factor was fixed to one to receive factor loadings from all three insight dimensions.

The model converged normally after 12 iterations. The latent insight factor loaded significantly onto all three variables: *Certainty* ($\lambda=.691$, $z=23.70$; $p<.001$), *Emotion* ($\lambda=.750$, $z=25.17$; $p<.001$) and *Suddenness* ($\lambda=.567$, $z=20.31$; $p<.001$) suggesting that those variables contribute significantly to the latent insight factor. Practical fit indices ($CFI= 1.0$; $RSME = 0.000$; $SRMR =0.000$) suggested a good fit of the model to the data. Note, the significant factor loadings did not change significantly when including all trials (i.e. also incorrectly identified Mooney images). In sum, those results justified combining the three insight dimensions into one single measure for further analyses.

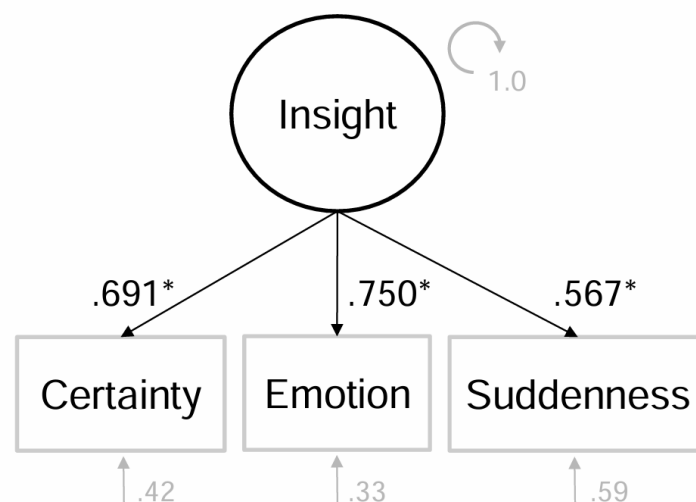


Figure S2. Measurement model for latent insight variable

Note. Asterisk = significant loadings at $p<.001$.

Estimating threshold for insight trials based on object recognition time

To eliminate trials that might solely involve object recognition and no insight, we excluded all trials with solution times faster than 1.5 seconds ($>5 \times SD$ of average object recognition time) from our analyses. The time for object recognition was informed by a control experiment, wherein a subset of gray-scaled images ($N=78$), the same set used to generate the Mooney images, was presented to 40 subjects (19 females, $M=27.3$ years, $SD= 5.6$ years) in an online test recruited via Mechanical Turk and a

German student platform of Humboldt University. Their objective was to press a solution button upon identifying the object, mirroring our Mooney identification task. Median correct object recognition time for these gray-scaled images was 753.2ms ($SD=140.1ms$) (see Fig. S3).

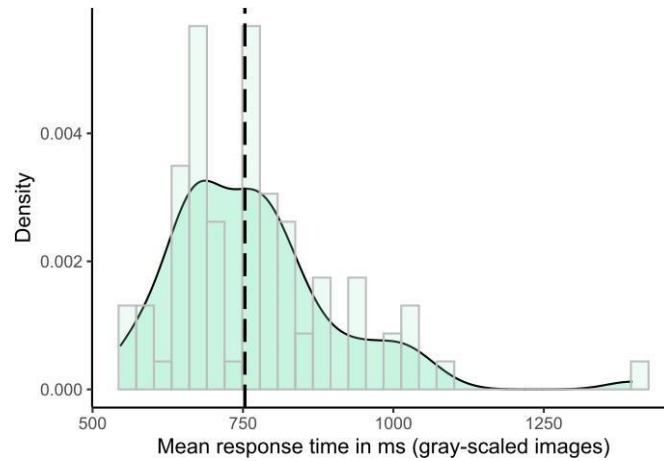


Figure S3. Mean response time for object recognition of gray-scaled images.

Note. This is a subset ($N=78$) of those images that were used to create the Mooney images. Dashed line represents median response time (753.2ms).

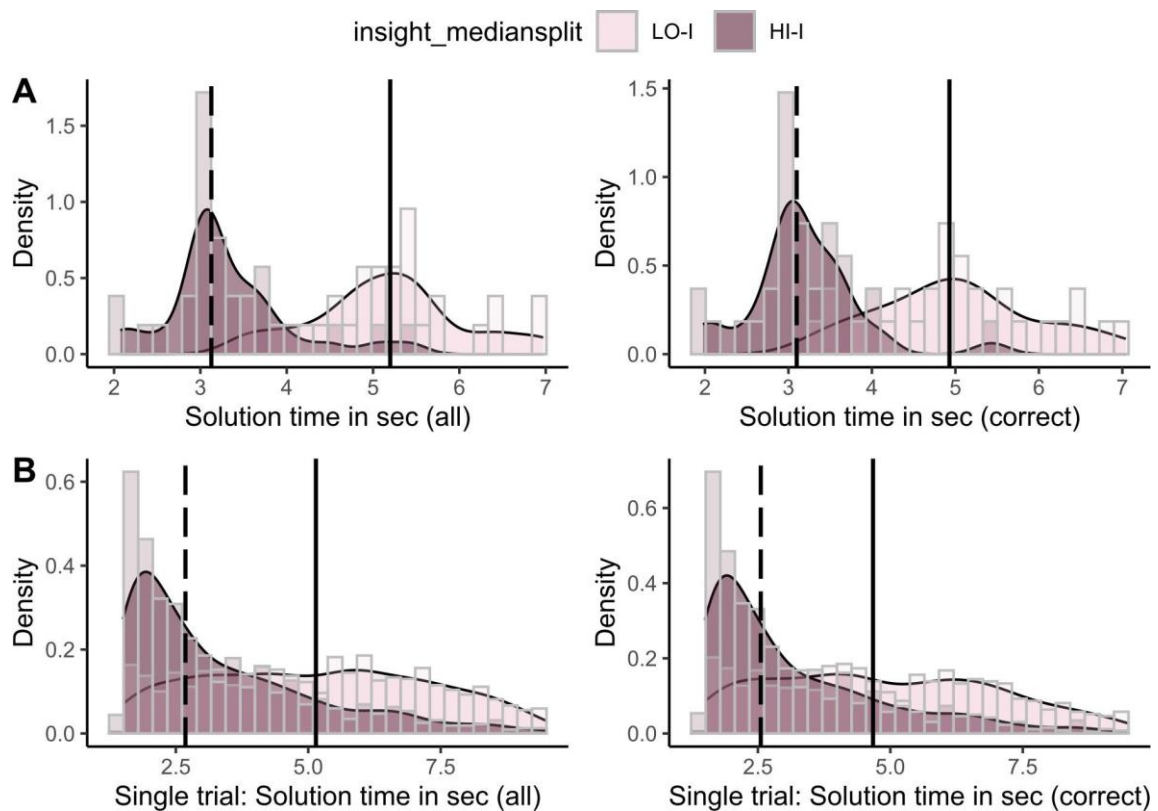


Figure S4. Response time distribution divided by high and low insight solutions

Note. Insight (sum of three 4-point Likert scales measuring its subdimensions: positive emotion, certainty, and suddenness) was divided into high (*HI-I*) and low (*LO-I*) insight via median split. (all) = all solved trials; (correct) = only correctly solved trials. Dashed line = median response time for *HI-I* trials;

Full line = median response time for *LO-I* trials. Panel A: Representation of the aggregated solution time distribution across each subject; Panel B: Illustration of the solution time distribution at the individual trial level.

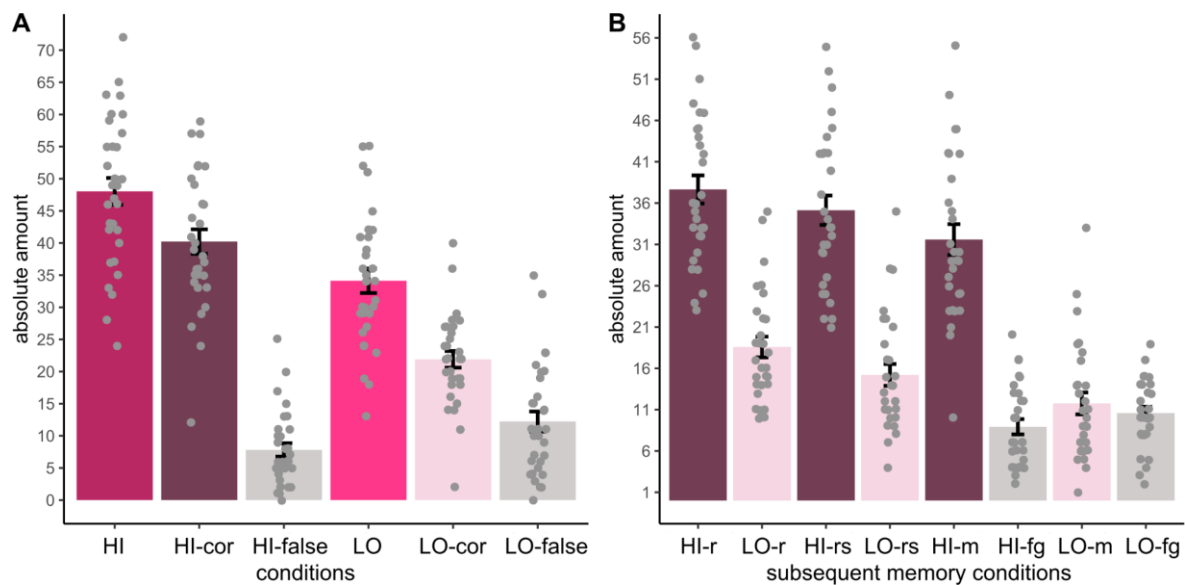


Figure S5. Amount of trials for each insight and insight memory condition

Note. The bars represent the average number of trials split by condition \pm SEM. **Panel A:** *HI*= high insight trials, *HI-cor*= correctly solved high insight trials, *HI-false* = incorrectly solved high insight trials, *LO* = low insight trials; *LO-cor* = correctly solved low insight trials; *LO-false*= incorrectly solved low insight trials; **Panel B:** *HI-r* = correctly solved high insight trials & later correctly recognized as having seen before; *LO-r* = correctly solved low insight trials & later correctly recognized as having seen before; *HI-rs*= correctly solved high insight trials & later correctly recalled as having solved before; *LO-rs* = correctly solved low insight trials & later correctly recalled as having solved before; *HI-m* = correctly solved and remembered high insight trials, *HI-fg* = correctly solved & forgotten insight trials, *LO-m* = correctly solved & later remembered low insight trials, *LO-fg*= correctly solved and later forgotten low insight trials.

Univariate fMRI results of VOTC RC areas

For control purposes, aiming to examine whether univariate brain activity in VOTC-RC regions could account for the observed multivariate differences related to insight, we explored the univariate brain activity in pFusG and iLOC in relation to insight. To investigate correctly solved insight trials, we performed the same single trial analyses as for amygdala and hippocampus controlling for run order, solution time and including random subject and item intercepts. The reported inferential statistics (χ^2) are based on nested model comparisons between two (general) linear mixed models where a baseline model is compared to a full model incorporating the independent variable of interest. For testing interaction effects, the baseline model additionally incorporates two independent variables, while the full model adds their interaction. Effect sizes—standardized beta estimates (β) for continuous outcomes—correspond to the regression estimate of the respective independent variable or interaction between two or three independent variables in the full model.

Regarding univariate effects of insight on VOTC-RC areas, there was no significant

overall insight effect ($Chi^2(1)=2.96$, $p=.085$, $\beta=-.02$, 95%CI[-.05, .00]) but an insight*ROI interaction ($Chi^2(1)=35.24$, $p<.001$, $\beta=.12$, 95%CI[.08, .16]) effect. Posthoc analyses revealed a negative association between insight and BOLD activity in iLOC ($z=-5.51$, $p<.001$, $\beta=-.85$, 95%CI[-1.16, -.55]) but no statistically significant evidence for an effect in pFusG ($z=1.27$, $p=.20$, $\beta=.15$, 95%CI[-.08, .39]) (see Fig.S6).

Moreover, we examined potential time-on-task influences on the BOLD signal in both VOTC-RC regions.⁵ To explore this, we replicated the single-trial analyses from above, excluding insight as a predictor. We compared two models: a full model including a solution time*ROI interaction, controlling for run order and incorporating random subject and item intercepts and a baseline model that is identical to the full model but without the interaction term. We found a significant solution time*ROI interaction ($Chi^2(1)=100.38$, $p<.001$, $\beta=-.20$, 95%CI[-.24, -.16]). Posthoc analyses revealed a positive association between solution time and BOLD activity in iLOC ($z=2.87$, $p=.004$, $\beta=.41$, 95%CI[.13, .69]) but a negative effect of solution time on BOLD activity in pFusG ($z=-12.11$, $p<.001$, $\beta=-1.28$, 95%CI[-1.49, -1.07]).

In summary, the inconsistent univariate effects observed in iLOC and pFusG in relation to insight and solution time make it implausible to attribute the observed consistent multivariate RC effects in both VOTC regions on univariate brain activity.

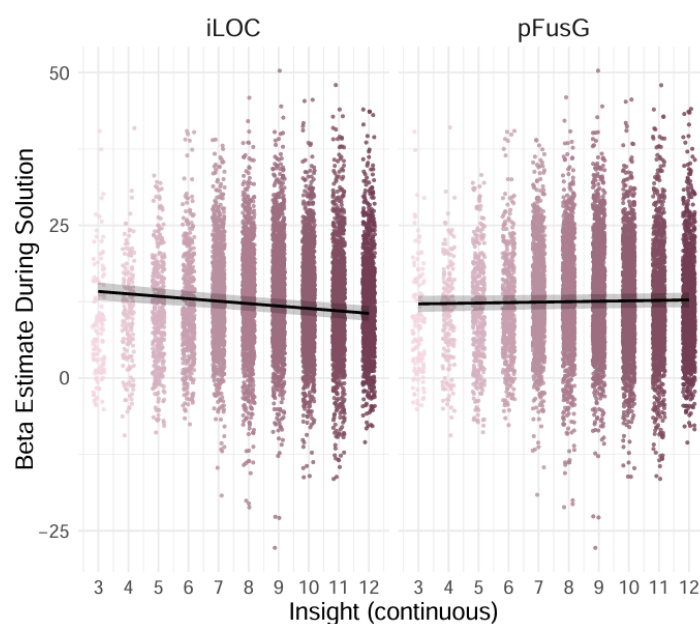


Figure S6. Relationship between univariate BOLD activity in RC VOTC areas and insight.

Note. Solid lines represent the regression weights of insight on univariate brain activity in iLOC ($z=-5.51$, $p<.001$, $\beta=-.40$, 95%CI[-.54, -.26]) and pFusG ($z=1.27$, $p=.20$, $\beta=.07$, 95%CI[-.04, .18]) (two-sided z-test, $n=9432$ samples), while the shaded region indicates the 95% confidence interval.

Further control analyses to account for the influence of solution time

We conducted additional control analyses to eliminate the potential influence of solution time as the sole factor responsible for the observed effects associated with insight.

Behavioral analysis: We adjusted solution time between *HI-I* and *LO-I* trials by excluding fast *HI-I* and slow *LO-I* trials until the time taken to solve both types of trials was no longer significantly different ($p > .20$). Subsequently, we repeated the mixed model analyses investigating memory as a function of insight (continuous) for correctly solved trials controlling for run (order) and random subject and item intercepts. Odds ratio (OR) and its 95% confidence interval (CI) for insight were non-parametrically bootstrapped (1000 resamples) via the *boot* function (v-1.3-30) in R. The results did not change significantly, insight (continuous) still positively predicted memory ($\chi^2(1)=63.83$, $p < .001$, OR: 1.576, 95%CI[1.41, 1.79]). This result further dismisses differences in solution time as potential explanation for the insight memory advantage.

We also conducted a preregistered control experiment (<https://aspredicted.org/xx7hv.pdf>, 11/28/2023) with originally 33 subjects and a final sample of $N=27$ subjects (70% females, $M=23.0$ years ($SD=4.1$ years)), recruited through the Humboldt University student online platform PESA. In this control experiment, participants replicated the same procedures as in the fMRI experiment, using the identical 120 Mooney images. However, the stimulus disappeared upon pressing the solution button, eliminating potential confounds related to differences in encoding time for the identified Mooney object. Second, the subjects typed their solution instead of selecting a response category which eliminated ambiguity regarding their solution. Participants performed the Mooney identification task in the lab, and five days later ($SD=0.3$ days), they underwent an online subsequent memory test. Six participants were excluded from further analyses due to producing too few correct hits (see preregistration). The data analysis followed the same procedures as for the fMRI sample, with the exception that there was no run (order) variable, as participants completed the experiment in a single run. A distribution of all trials is depicted in Fig. S7-A.

On average, participants found a solution in 45.5% ($SD=15\%$) for all presented Mooney images (see Fig. S7-A). The solution was correct in 60.2% ($SD=17\%$) of these cases. Of all correctly identified Mooney images, 57% were solved with *HI-I* and 43% with *LO-I* trials ($SD=12\%$). Consistent with previous results and the fMRI sample, this *insight-accuracy* effect was significant ($\chi^2(1)=166.0$, $p < .001$, OR=1.55, 95%CI[2.51, 3.64]).^{6,7} The median response time for correct solutions was 4.3sec ($SD=0.77$ sec) where participants were faster during *HI-I* (3.6sec, $SD=0.67$ sec) than *LO-I* solutions (5.3sec, $SD=1.0$ sec). Similar to the fMRI sample, insight significantly predicted response time ($\chi^2(1)=210.42$, $p < .001$, $\beta = -.21$, 95%CI[-.24, -.19]). Only accurate responses and response time as covariate of no interest were included into all subsequent memory analyses. Participants recognized 49.0% ($SD=16.5\%$) of the 120 presented Mooney images, regardless of whether they had correctly identified them, with a false alarm rate of 16.3% ($SD=12.1\%$). Out of all presented Mooney images, participants correctly solved *and* subsequently correctly recognized 18% ($SD=9\%$) and they correctly solved and subsequently correctly recalled having solved them in 15% ($SD=8\%$). Finally, out of all presented Mooney images, participants correctly solved and subsequently correctly identified them in 15% ($SD=8\%$) (see Fig.S7-B for an overview of a distribution across all conditions). The insight factor (*HI-I*, *LO-I* and

unsolved trials) significantly predicted variance in subsequent memory ($\chi^2(2)=113.2$, $p<.001$, OR (linear trend for not solved < LO-I < HI-I) = 7.42, 95%CI[4.93, 11.16]; see Fig.S7-C). Importantly, when exchanging the binarized insight variable in the mixed model with the continuous insight variable additionally controlling for solution time (excluding unsolved trials), insight still significantly predicted subsequent memory ($\chi^2(1)=25.41$, $p<.001$, OR=1.28, 95%CI[1.16, 1.40]) (see Fig.S7-D).

To summarize, the behavioral control experiment replicated the insight memory advantage observed in the fMRI sample. These findings dismiss longer encoding time of Mooney objects as a potential explanation for the observed insight memory advantage.

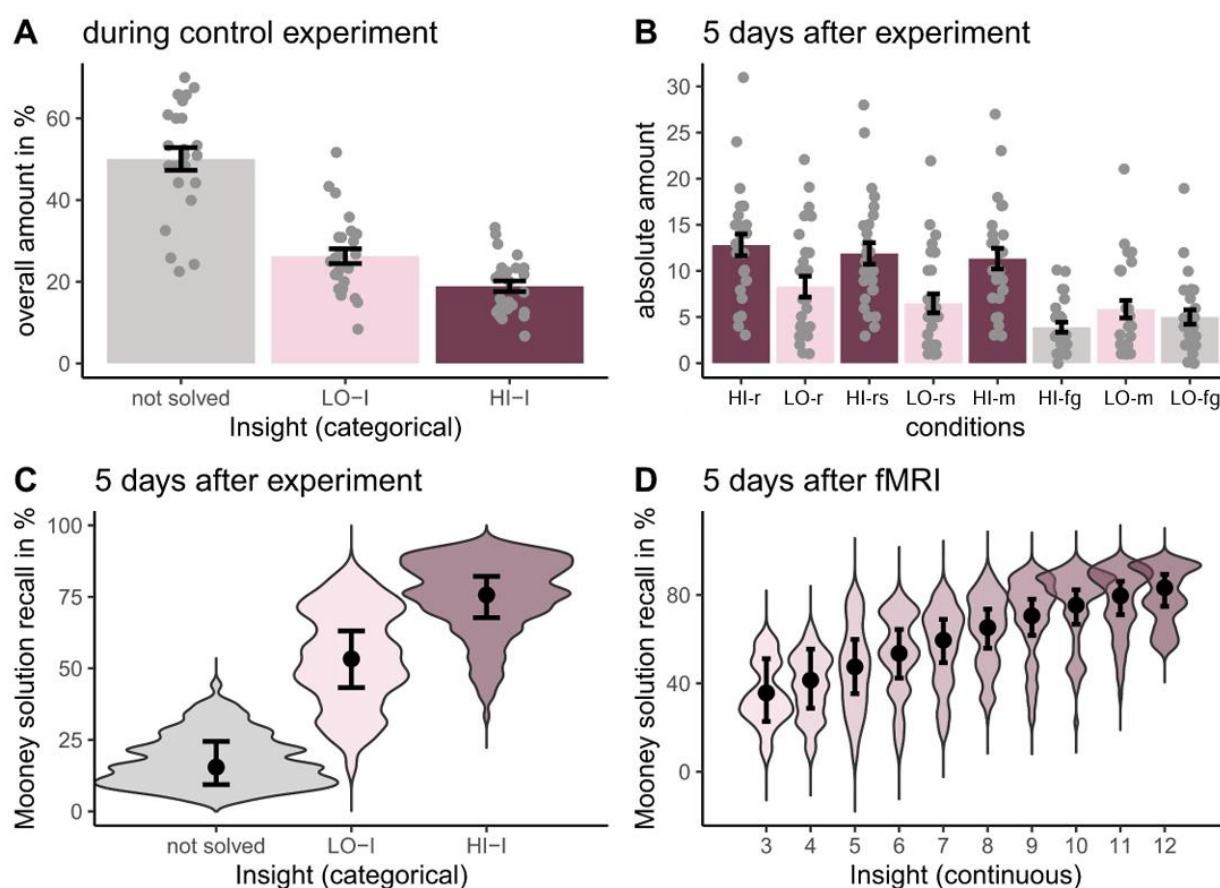


Figure S7. Behavioral Insight Memory Effect - control experiment.

Note. **Panel A.** Overall proportion of all trials per condition error bars represent SEM; $n=27$ samples; LO-I = all low insight trials; HI-I = all high insight trials; not solved = unsolved trials. **Panel B:** (remembered = Mooney object correctly identified) HI-r = correctly solved high insight trials & later correctly recognized as having seen before; LO-r = correctly solved low insight trials & later correctly recognized as having seen before; HI-rs = correctly solved high insight trials & later correctly recalled as having solved before; LO-rs = correctly solved low insight trials & later correctly recalled as having solved before; HI-m = correctly solved and remembered high insight trials, HI-fg = correctly solved & forgotten insight trials, LO-m = correctly solved & later remembered low insight trials, LO-fg = correctly solved and later forgotten low insight trials; error bars represent SEM ($n=27$ samples); **Panel C.** Behavioural insight memory effect ($\chi^2(2)=113.2$, $p<.001$, OR=7.42, 95%CI[4.93,11.16]) based on a two-sided likelihood ratio test comparing a model with and without (categorical) insight; HI-I = correctly solved high insight trials; LO-I = correctly solved low insight trials; dots are marginal mean percentages

to correctly recall the previously identified Mooney image in the respective condition. Error bars represent 95% confidence intervals. **Panel D.** Relationship between continuous insight variable (3-12) and memory controlled for solution time ($\chi^2(1)=25.41$, $p<.001$, $OR=1.28$, $95\%CI[1.16, 1.40]$) based on a two-sided likelihood ratio test comparing a model with and without (continuous) insight; black dots represent marginal mean percentages per insight value (3-12) to correctly recall the previously identified Mooney image ($n=704$ samples).

Univariate BOLD analysis: Similar to the behavioural analysis, we examined the impact of removing fast HI-I and slow LO-I trials on BOLD activity (fast HI-I and slow LO-I trials were removed until the time taken to solve both types of trials was no longer significantly different [$p>.20$]). Subsequently, we conducted repeated single-trial mixed-model analyses investigating BOLD activity in amygdala and hippocampus as a function of a) insight (continuous) as well as b) insight*memory controlling for run (order) and random subject and item intercepts. To enhance statistical robustness, the p-value was obtained through permutation tests ($n = 999$) by comparing nested random effects models using the *predictmeans* package (v.1.1.0) in R.⁸ A) The model including insight to predict BOLD activity during solution performed significantly better than the model without insight ($\chi^2(1)=32.43$, $p<.001$, $\beta=.09$, $95\%CI[.06, .12]$). The model with the insight*ROI interaction was not statistically significantly better than the baseline model without this interaction term ($\chi^2(2)=2.33$, $p=.31$). B) similarly for insight, there was a significant insight*memory interaction ($\chi^2(1)=8.09$, $p=.008$, $\beta=.09$, $95\%CI[.03, .14]$) predicting BOLD activity in all three ROIs while there was no statistically significant evidence for a three-way interaction (ROI*insight*memory: $\chi^2(6)=5.79$, $p=.45$). A Post-hoc analysis revealed that the slope for insight during remembered trials was significantly positive ($z=6.23$, $p<.001$, $\beta=.17$, $95\%CI[.12, .22]$) while it was not statistically significantly different from zero during forgotten trials ($z=1.65$, $p=.10$, $\beta=.05$, $95\%CI[-.01, .12]$).

Multivariate BOLD analysis: Making adjustments for solution time in these correlational measures by excluding fast HI-I and slow LO-I trials is not plausible, as each trial represents a correlation of the multivoxel pattern of the current trial in a given ROI with the pattern of every other trial, whether solved fast or slow. However, to account for the potential impact of BOLD activity on the relationship between insight and multivariate measures due to potential time-on-task effects, we conducted a repetition of the aforementioned multivariate RC analyses. This time, we directly incorporated BOLD activity during solution as a covariate of no interest in exchange for solution time additionally controlling for run (order) and random subject and item intercepts. Note, to enhance statistical robustness for the insight-memory analyses, p-values were obtained through permutation tests ($n=999$).⁸

The model including insight performed significantly better than the baseline model without insight a predictor for multivoxel pattern similarity (first RC measure) from pre to post solution ($\chi^2(1)=313.08$, $p<.001$, $\beta=-.14$, $95\%CI[-.15, -.12]$) for both VOTC RC regions also controlling for run (order) and BOLD activity during solution. Post-hoc tests showed that pFusG ($\chi^2(1)=211.61$, $p<.001$, $\beta=-.22$, $95\%CI[-.25, -.19]$) as well as iLOC ($\chi^2(1)=155.70$, $p<.001$, $\beta=-.25$, $95\%CI[-.29, -.21]$) reduced their multivoxel pattern similarity from pre to post solution with increasing insight.

Furthermore, the model including the insight*time(post solution) interaction predicting representational strength (AlexNet, second RC measure) performed significantly better ($\chi^2(1)=15.19$, $p<.001$, $\beta=.06$, 95%CI[.03, .09]) than the model without this interaction term for both RC regions (pFusG, iLOC) when controlling for BOLD activity during solution next to run (order) and random subject and item intercepts. The model with the additional three-way interaction (insight*time*ROI) was not statistically significantly better than the model without it ($\chi^2(3)=1.98$, $p=.58$). Post-hoc analyses revealed that insight negatively predicted representational strength during stimulus onset ($z=-2.27$, $p=.023$) but positively during solution ($z=2.79$, $p=.005$) over both ROIs. Regarding insight-related memory, there was no statistically significant interaction between insight, memory(remembered), and time (post solution) ($\chi^2(1)=0.71$, $p\text{-perm}=.40$, $\beta=-.03$, 95%CI[-.10, .04]). In contrast, the interaction between memory(remembered) and time (post solution) remained significant ($\chi^2(1)=50.85$, $p\text{-perm}=.001$, $\beta=.24$, 95%CI[.17, .30]).

When using Word2Vec as a conceptual model to compute representational strength, the model with the insight*time(post solution) interaction predicting representational strength (second RC measure) remained performed significantly better than the model without this two-way interaction ($\chi^2(1)=16.12$, $p<.001$, $\beta=.06$, 95%CI[.03, .09]) for both RC regions (pFusG, iLOC) controlling for BOLD activity during solution next to run (order) and random subject and item intercepts. The model with the insight*time*ROI interaction did not perform better than the model without this three-way interaction ($\chi^2(3)=1.57$, $p=.66$). Post-hoc analyses revealed that insight positively predicted representational strength during solution ($z=3.67$, $p<.001$, $\beta=2.74^{-3}$) while there was no statistically significant evidence for a difference in representational strength as a function of insight during stimulus onset ($z=-1.55$, $p=.122$, $\beta=-1.15^{-3}$). Furthermore, there was a significant positive insight*memory(remembered)*time(post solution) interaction ($\chi^2(1)=10.46$, $p\text{-perm}=.004$, $\beta=.09$, 95%CI[.02, .15]) predicting representational strength over both brain regions (iLOC, pFusG) while this interaction was not statistically significantly different between both ROIs ($\chi^2(7)=6.63$, $p\text{-perm}=.46$).

In summary, the outcomes remained largely consistent when incorporating BOLD activity during the solution as a covariate of no interest instead of solution time in the main analyses outlined in the results section.

Exploratory whole-brain analysis

For exploratory purposes, we also performed a whole-brain analysis to examine whether activity in other brain regions might be parametrically modulated by the intensity of the Aha! experience.

First-Level Analysis. We utilized the beta values previously estimated during the first-level analysis (as detailed in the ROI analysis, see Method section). For each participant, simple contrast t-images were generated from the beta weights associated with the onset regressors.

Second-Level Analysis. The contrast images representing the parametric modulation of the Aha! experience during the solution phase (button press) were analyzed using SPM's one-sample t-test. Multiple comparisons were corrected by applying a voxel-level threshold of $p < .001$ ⁹ and a cluster-level threshold of $p < .05$ (family-wise error corrected), and with a height threshold of $t = 3.40$. All anatomical regions were identified using the AAL3 atlas¹⁰, based on the percentage of voxels within an activated cluster from the one-sample t-test that intersected the respective anatomical region.

Whole-Brain Analysis Results: The intensity of the Aha! experience showed a positive correlation with three distinct clusters (extent threshold $k = 132$ voxels; see Fig. S8). The primary cluster (size: 1,081 voxels, xyz[22, -8, -20]) was located around bilateral hippocampus (covers 32% of this cluster) and extended to bilateral amygdala (11.5%), parahippocampal gyrus (12.5%), ventral striatum (7.5%), olfactory bulb (5.5%), and bilateral putamen (14%). The fact that the largest cluster appeared around the hippocampus and amygdala validates our previous hypothesis that these regions are particularly associated with insight's evaluative component. The other two clusters were medially situated, with an anterior cluster (size: 341 voxels, xyz[-2, 34, 8]) encompassing the anterior cingulate cortex (70%) and bilateral medial orbital frontal cortex (14%). The ACC has been frequently linked to insight^{11,12}, the OFC has also been reported as part of the reward network during insight.¹³ The posterior cluster (size: 458 voxels, xyz[-14, -52, 36]) included the bilateral precuneus (70%), bilateral posterior (15%) and middle (7.5%) cingulate cortex, and the left cuneus (2.5%). The involvement of the precuneus, which is associated with semantic retrieval¹⁴, is plausible as participants were likely retrieving the semantic content of the object at that moment.

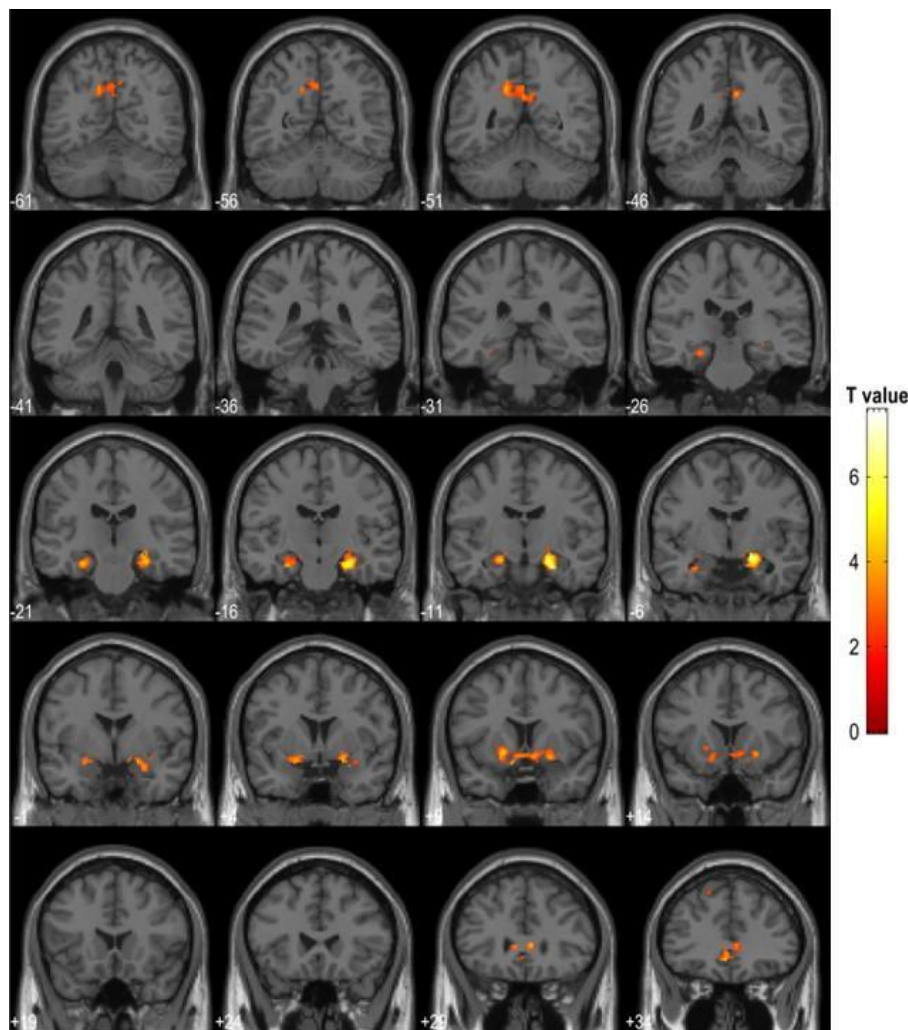


Figure S8. Whole-Brain Activation positively correlated with intensity of Aha! experience

Note. Threshold for visualisation: voxel-based threshold $p < .001$, cluster-threshold $p\text{-FWE} < .05$; $n = 31$ samples.

Table S1.

Individual Connections from "solution network" for HI-I > LO-I

ROI1	ROI2	T(30)	p (uncorrected)
iLOC l	mITG l	4.32	0.000079
iLOC r	mITG l	3.46	0.000822
pITG l	mITG l	3.09	0.002136
pFusG r	mITG l	3.02	0.002557
iLOC r	aFusG r	3.01	0.002602
iLOC l	aFusG r	2.88	0.003656
iLOC r	mITG r	2.76	0.004900
iLOC l	mITG r	2.66	0.006275
pFusG l	mITG l	2.65	0.006405
aHC r	aITG l	2.61	0.007060
pFusG l	pITG r	2.54	0.008280
pHC l	aITG l	2.33	0.013467

iLOC l	pFusG l	2.32	0.013530
pFusG l	mITG r	2.27	0.015252
pFusG r	pITG r	2.25	0.015915
aHC r	Amygdala l	2.23	0.016675
pFusG r	aHC l	2.21	0.017485
pFusG l	aITG l	2.15	0.019711
pITG r	mITG l	2.11	0.021769
pFusG r	aITG l	2.04	0.024871
pFusG r	aFusG r	2.01	0.026689
iLOC r	Amygdala r	2.01	0.026762
mITG r	aITG l	1.97	0.029360
iLOC l	Amygdala l	1.95	0.030032
iLOC l	Amygdala r	1.91	0.032553
aHC l	aITG l	1.88	0.034908
aFusG r	Amygdala r	1.85	0.036759
aFusG l	mITG l	1.82	0.039460
pFusG r	Amygdala r	1.82	0.039621
iLOC r	aHC r	1.80	0.040669

Note. Individual functional connections from a 18-ROIs "solution" network containing a/pFusG, iLOC, a/m/p ITG, Amygdala, aHC and pHC (r = right; l = left) averaged across 31 subjects (one-sample T-test statistics (one-sided), p-values are uncorrected for multiple comparison).

Graph measures for characterizing insight-related more efficient information integration within the solution network

Degree serves as a reliable indicator of centrality for both networks and ROIs, capturing the level of local interconnectedness within a graph or the entire network. At the ROI level, *Degree* quantifies the edges connected to/from each ROI, while the network degree calculates the mean degree across all graph nodes.¹⁵

Functional integration within the brain denotes its capacity to rapidly combine specialised information from distributed brain regions, thus illustrating the fluidity of inter-regional communication.¹⁶ Metrics indicating functional integration or the extent of overall interconnectedness in the network comprise *Average path distance* and *Global efficiency*. The *Average path distance* between pairs of ROIs in a graph refers to the minimum connections to navigate between them along an optimal route. Consequently, the node-specific average path distance pertains to the mean path distance between the node and all other nodes in the connected subgraph. *Global efficiency* is calculated as the average of inverse-distances between a given ROI and all other ROIs within the same graph.¹⁵

Functional segregation within the brain refers to its capacity for specialized processing to unfold within densely interconnected clusters of brain regions.¹⁶ Indicators of functional segregation or localized integration encompass the *Clustering coefficient* and *local efficiency*. The *Clustering coefficient* is delineated as the ratio of interconnected edges in the nearby sub-graph corresponding to each ROI. Node-specific *Local*

efficiency is defined as the Global efficiency within the neighboring sub-graph of the given ROI.¹⁵

Table S2.

Mathematical definitions of the used graph measures as implemented in CONN

Function	Graph Measure	ROI-Level	Network-Level
Centrality	Degree (d)	$d_i = \sum_j A_{i,j}$	$d = \frac{\sum_i d_i}{N}$
Function integration	Average path length (L)	$L_i = \frac{\sum_{j \in \Omega_i} D_{i,j}}{N_i - 1}$	$L = \frac{\sum_i L_i}{N}$
	Global Efficiency (GE)	$GE_i = \frac{\sum_{j \neq i} 1/D_{i,j}}{N - 1}$	$GL = \frac{\sum_i GE_i}{N}$
Functional segregation	Clustering Coefficient (CC)	$CC_i = \frac{\sum_{j,k \in \Gamma_i} A_{j,k}(i)}{d_i(d_i - 1)}$	$CC = \frac{\sum_i CC_i}{N}$
	Local Efficiency (LE)	$LE_i = \frac{\sum_{j \neq k \in \Gamma_i} 1/D_{j,k}^{(i)}}{d_i(d_i - 1)}$	$LE = \frac{\sum_i LE_i}{N}$

Note. A = adjacency matrix; N = total number of nodes in a graph; d_i = degree of a graph at each node; D = shortest-path distance matrix, D_x is the shortest-path distance matrix within the neighboring sub-graph at each node, characterized by all nodes neighboring this node and all existing edges among them

Relationship between insight-related efficient information integration and subsequent memory

Finally, for reasons of completeness, we further explored whether different graph measures in the pre-defined *solution network* also predicted insight-related better memory. For this, we repeated FC and graph theoretical analyses but instead of a simple *HI-I>LO-I* contrast, we now calculated an insight (*HI-I* vs *LO-I*) x memory (remember vs. forgotten) interaction with four conditions: 1) high insight - remembered trials [*HI-I-Rem*], 2) high insight - forgotten trials [*HI-I-Forg*], 3) low insight - remembered trials [*LO-I-Rem*] and 4) low insight - forgotten trials [*LO-I-Forg*] to construct the adjacency matrices. To calculate an interaction, we used the following contrast $1*HI-I-Rem - 1*HI-I-Forg - 1*LO-I-Rem + 1*LO-I-Forg$ for subsequent FC and graph theoretical analyses. However, please note that due to the low trial count and resulting reduced statistical power, these results should be considered only exploratory.

Functional connectivity. The *solution network* did not survive multiple comparisons at the cluster-threshold for an insight x memory interaction.

Graph theoretical measures. Global efficiency was significantly increased for this interaction ($t(30)=1.74$, $p<.05$ (one-tailed [directional hypothesis]), $\beta=.03$) at the *solution network*-level and also right iLOC ($t(30)=1.70$, $p\text{-uncor}<.05$, $\beta=.03$) next to left ($t(30)=3.59$, $p\text{-FDR}<.05$, $\beta=.14$) and right aITG ($t(30)=1.94$, $p\text{-FDR}<.05$, $\beta=.06$) exhibited significantly more global efficiency compared to the other ROIs in the network. However, the insight*memory interaction in the *solution network* failed to reach statistical significance for increased local efficiency ($t(22)=1.45$, $p=.08$, $\beta=.03$). Furthermore, no individual ROI survived multiple comparisons for local efficiency. Neither degree ($p>.73$), average pathlength ($p>.97$) nor clustering coefficient ($p>.15$) showed significant effects in the hypothesized direction.

In sum, the results from FC and graph theoretical measures provide no consistent evidence although some graph measures (global efficiency) suggest an association between insight-related efficient information integration and better subsequent memory.

Supplement References

1. Rosseel, Y. **lavaan** : An R Package for Structural Equation Modeling. *J. Stat. Softw.* **48**, (2012).
2. Hu, L. & Bentler, P. M. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychol. Methods* **3**, 424–453 (1998).
3. Hu, L. & Bentler, P. M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscip. J.* **6**, 1–55 (1999).
4. Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods Psychol. Res.* **8**, 23–74 (2003).
5. Yarkoni, T., Barch, D. M., Gray, J. R., Conturo, T. E. & Braver, T. S. BOLD Correlates of Trial-by-Trial Reaction Time Variability in Gray and White Matter: A Multi-Study fMRI Analysis. *PLoS ONE* **4**, e4257 (2009).
6. Danek, A. H. & Salvi, C. Moment of truth: Why Aha! experiences are correct. *J. Creat. Behav.* **54**, 484–486 (2020).
7. Becker, M., Wiedemann, G. & Kühn, S. Quantifying insightful problem solving: A modified compound remote associates paradigm using lexical priming to parametrically modulate

- different sources of task difficulty. *Psychol. Res.* **84**, 528–545 (2020).
8. Lee, O. E. & Braun, T. M. Permutation tests for random effects in linear mixed models. *Biometrics* **68**, 486–493 (2012).
 9. Eklund, A., Nichols, T. E. & Knutsson, H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci.* **113**, 7900–7905 (2016).
 10. Rolls, E. T., Huang, C.-C., Lin, C.-P., Feng, J. & Joliot, M. Automated anatomical labelling atlas 3. *NeuroImage* **206**, 116189 (2020).
 11. Dietrich, A. & Kanso, R. A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychol. Bull.* **136**, 822 (2010).
 12. Becker, M., Cabeza, R. & Kizilirmak, J. M. A cognitive neuroscience perspective on insight as a memory process: Searching for the solution. in *The Routledge international handbook of creative cognition* 491–510 (Routledge, 2023).
 13. Oh, Y., Chesebrough, C., Erickson, B., Zhang, F. & Kounios, J. An insight-related neural reward signal. *NeuroImage* **214**, 116757 (2020).
 14. Flanagan, V. L. *et al.* The precuneus as a central node in declarative memory retrieval. *Cereb. Cortex* **33**, 5981–5990 (2023).
 15. Whitfield-Gabrieli, S. & Nieto-Castanon, A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.* **2**, 125–141 (2012).
 16. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).