doi:10.1038/nature09510

Verification of the decoding

As another measure of selectivity, we decoded the target on a trial by trial basis purely by the firing rate of the four recorded units. For 92% of successful fading trials, we could identify the target unit purely by testing whether the average firing rate of one out of four units we recorded increased above its baseline in all 8 fading trials for a given image. This clear change in neuronal firing rate based on the internal state of the subject indicates that the neuronal feedback was given by the subject's thought and not by the external stimulus.

Response characteristics

Testing for interaction between the units using temporal and within-trial cross-correlations, either by pairing all units whose preferred stimulus was the target and those whose preferred stimulus was the distractor within or across all regions, or during a 100ms window within a trial, did not reveal any statistically significant temporal lag. This is likely due to the low overall spike count per bin for our analyses.

Out of image sets ranging from 93 to 136 images (110.8 ± 14.9, different set size for each session), 133 units responded to 5.63 images on average, 4.2% of all images shown. The average baseline firing rate, computed during the control presentations, was 4.2Hz, with a mean standard deviation of 2.6Hz. The average increase in firing rate above baseline (in standard deviation units) during all fading trials where the unit's preferred stimulus was the target was 3.72 (compared to increased activity of at least 5 standard deviations above the baseline for visual presentation during the screening experiment). This is calculated by subtracting off each unit's baseline firing rate and normalizing by the standard deviation of this baseline. During successful trials, the average firing rate of a MTL unit whose preferred stimulus was the focus of the fading experiment was 11.1 ± 7.4Hz. The average firing rate of the unit whose preferred stimulus was the distractor, during successful fading trials, was 1.9 ± 2.6Hz, corresponding to a normalized decrease of 0.59.

Single and multi-units

While performance was based on spike detection prior to spike discrimination (to maximize processing speed for real-time performance), we carried out a *post-hoc* analysis to distinguish single- from multi-units. Using the *wave_clus* spike sorting algorithm[1], we sorted the data and identified 14 out of 72 units as single-units. We allowed up to 10% difference in the number of spikes between the original unit used and the sorted single-unit for inclusion in the *post-hoc* analysis. For example, if a unit which was used in the experiment was regarded a multi-unit and had, say, 4,500 spikes throughout the experiment, but after sorting was identified as a single-unit with 4,200 spikes (300 additional spikes were either considered artifacts or coming from a different unit) then the single-unit was included in the following analysis. Using these single-units alone, we considered the 112 trials (out of the 864) where the competition included at least one single-unit, and compared their performance against 752 multi-unit trials. The overall performance of the 58 pairs of multi-units in the experiment was 69.0% (519 trials won, 171 lost, 62 timeouts), while the performance of units where at least one was a single-unit was 68.8% (77 wins, 27 losses and 8 timeouts). The difference is not significant ($p = 0.1$), suggesting that subjects did equally well in controlling either types of units.

Comparing the average level of control of single and multi-units using the TDC metric showed no significant difference ($p = 0.24$, t-test), with the average TDC for single-units being 0.36 ± 0.28 and for multi-units being (0.43 ± 0.30). Subjects cannot clearly control single-units better than multi-units or vice versa.

Performance in the very first trials

When the first two trials from each subject's block were considered as *de facto* training trials, the performance of the remaining trials improved by 3.8% to 72.8%.

Learning to delay failure

Subjects could control which one of two images dominated within a single session, that is, within a few minutes. As success involves a combination of suppressing the distractor image and enhancing the target, we monitored the degree of learning in two complementary ways, focusing on whether subjects took longer to fail and/or became faster at winning. Supplementary Figure 5 plots the time-to-fail for two subjects. For the first subject

(Supplementary Fig. 5a), in a competition pitting a picture of the actor Denzel Washington against a picture of a building in Las Vegas the subject was familiar with, he failed in all 8 trials. Strikingly, the time-to-fail lengthened over the 8 trials. Considering all 8 blocks where all subjects failed in all trials of a single target, the time-to-fail increased as subjects gained more experience (Supplementary Fig. 5c; $p < 10^{-5}$, Spearman's rank correlation), with an average slope of 0.89 ± 0.21 s/trial. That is, after each failed trial, subjects took 0.89s longer to fail on the next trial. Time-to-fail as compared to the previous trial increased in 99% of successive failed trials (and not just in those with 8 consecutive failed trials). In the 12 blocks where subjects achieved 100% visibility of the target in all 8 consecutive trials, no timing difference was apparent. The difference between consecutive successful trials was -0.14 ± 0.67s (n.s; p = 0.42, Spearman's rank correlation). This might be due to a floor effect, *i.e.*, the time-to-success on the first trial is already close to the theoretical minimum of 1s (as the visibility is updated by 5% every 100ms, starting from 50%). We conclude that successful manipulation of neural firing can be achieved rapidly, and often within the first trial. In those cases where subjects failed to control neuronal firing, they at least learned to delay failure. Supplementary Fig. 5b shows an additional example from a different subject.

During sham trials, the mean increase in time between successive failed trials -0.09 ± -0.78s (p > 0.20) and mean speedup for successful trials -0.13 ± 0.63s (p > 0.40). We analyzed neuronal control during sham trials by binning each 100ms step for each trial by the output of the decoder. No consistent trend was seen in the firing rates of either target or distractor units during sham feedback.

Are there any long-lasting effects of feedback on the excitability of the MTL neurons? That is, do those neurons whose firing rate was up- or down-regulated by subject's thoughts retain any chronic changes in their responsivity? We used five criteria to test for changes in neuronal activity in the control presentations before and after the game: latency, duration, peak firing rate, mean firing rate, and time-of-peak activity of the individual units. No significant change was seen in any of these parameters. This suggests that either the feedback had no lasting effect on the neurons, or any sustained effect is not apparent when subjects are exposed to the images in passive viewing during our control presentation procedure. The absence of any explicit performance-based reward might also play a role here.

Imagery is capable of overriding distracting sensory input

We directly compared vision and imagery in the situations at which the two are pitted. Out of the 235 (27.2%) trials where at some stage of the trial the distractor had a higher visibility than the target, the subject was able to eventually win 71.7% of those trials. That is, the composite image shifted back towards the target despite the distractor being more visible than the target. If fading is entirely controlled by bottom-up, retinal input, these trials would be expected to end in a loss. To test the significance of these winning trials, we bootstrapped trials based on the subjects' proportions of 100ms bin that shifted toward or away from the target image (see Method) and compared them to those trials where the distractor was dominant. This demonstrates that the majority (88.1%) of such cases would have ended in failure, instead of actually being successful ($p < 0.01$, Wilcoxon rank-sum, shuffling trials based on the proportions starting with the *a priori* bias towards the distractor).

Testing for low-level strategies to control the feedback

Two additional control subjects had four responsive units during the morning's screening session, each unit responding selectively to a different image. However, during the afternoon session, none of those units retained their selectivity (as assayed during the control presentation). This could have been due to electrode movement, inflammation, plasticity, a seizure which occurred between the two sessions, or other factors beyond our control. The performance of these two subjects during the fading was at chance (21% and 16%), with 22 trials ending in a timeout for both. This is the highest percentage of timeouts over all subjects. For comparison, the average number of timeouts for the other 12 subjects is 7.04 ± 3.60 trials. What these two subjects demonstrate is that unless the units that carry out the decoding fire selectively to specific images, the subjects are unable to perform the task even though they are trying to enhance the target and/or suppress the distractor. Motor strategies to control these units should have been as easy for subjects to discover here as for units whose visual selectivity remains constant.

It is unlikely that subjects adapted an explicit motor strategy to control their units. To do so, each subject would have to: (i) discover the relevant motor strategies within a few trials. This is contradicted by some of our subjects that show effective neuronal control on their first trials using – as they reported to us – an explicit cognitive strategy. There was simply no time to try out different motor strategies. Consider Fig. 2: even though this is the first time ever the subject was asked to control her neurons, she was successful on all 8 trials; (ii) adopt different strategies to accurately control the firing rate of the four distinct units (e.g., right-ward saccade for the unit selective to image A, batting an eye-lid for unit B and so on); (iii) use them accurately without being noticeable either to us experimenters nor to the subject him- or herself. None of our subjects ever reported having used any overt motor strategy. We neither encouraged nor discouraged such strategies and therefore subjects had no reason not to report them; (iv) Finally, subjects should have been able to use the same motor strategy during sham trials, which should have resulted in a much higher performance during those.

Specificity of feedback and selective units

Is the specificity of our results confined to the individual units used by the decoding or is it due to a generalized change of firing rate of any suitable selected subset of units? We pooled all units not used in the experiment to ascertain their performance during feedback. Out of 779 units not used in the fading experiment (due to them not being responsive to the images used), we picked 4 units at a time and generated 32 trials sessions based on the activity of those, generating fictive fading sessions. We repeated these quadruplets selection 100 times for each subject. None of these other units' performance was significant when tested against 1000 Monte-Carlo realizations. The average performance of the 100 sessions x 12 subjects, using this method, was 20.6%, significantly below the performance of selective units ($p = 0.01$, Wilcoxon rank-sum). That is, high performance is confined to the units whose preferred images are shown rather than to a random pool of 4 units.

Additionally, to show that the modulation of activity is exclusive to the selective units used rather than to nearby units or an entire region, we tested the performance of neighboring units in the same task. That is, if the units used in the experiment were from the right amygdala and the left parahippocampal cortex, we selected four units in these regions that were not the ones used in the experiment and computed new fictive fading sessions based on feeding the activity of these four units into the decoder as a control. For each subject, we selected 100 quadruplets of units to replace the original selective ones, and for each we tested the performance in comparison to 1000 bootstrapped trials. Performance was not significant. This suggests that while subjects are able to control and modulate the activity of four units selective to specific images, this control is not generalized to an entire region.

While the feedback to the subject was based on the activity of four selective units, it is extremely likely that there were many more units in the subject's brain that responded to the same image (we estimated the size of this population using Bayesian reasoning in [2]). Accordingly, if the target unit was responsive to a picture of Johnny Cash, we looked for the rare case where we could locate one or more additional units responsive to Johnny Cash's image that were not used to control the fading. The subject most likely activated a large pool of neurons selective to 'Johnny Cash' even though the feedback was only based on just one such unit. We identified 8 such units in a total of 7 subjects. In a *post-hoc* analysis where we used these 'sister' units instead of the original target unit, the performance was 52.3%. For illustration, the average chance performance (as seen in the rightmost red bar in Fig. 3a) was 35.7%. The results are significant when compared, additionally, to chance calculated using the bootstrapping method ($p < 0.001$, Wilcoxon rank-sum), with chance being on average 37.0%. The performance of these 'sister' neurons suggests that, indeed, a population of neurons responding selectively to the target became activated during fading.

Image saliency

When images were first presented to subjects in the beginning of a trial, each image's transparency was set to 50%. However, some images might be more dominant than others even when balanced in such a way because of their coloring or content. This could affect the initial direction of a trial, as the dominant image might draw more attention from the subject, causing an initial movement towards it. While it is hard to tell which image draws the attention of our subjects more objectively, we tested all images pairs viewed by our subjects with the standard Itti-Koch saliency model[3]. We computed a saliency map for the first 3 most salient locations on the 50%/50% hybrid image. This allowed us to verify that no pair had in any of the 3 most salient locations a patch of any of the images that was more visible than the other (e.g. for the combination of Marilyn Monroe and Josh Brolin images shown in Fig. 2, when computing the standard saliency map model, no patch either from the Brolin image or from the Monroe image was drawing attention in the first 3 fixations). This suggests that none of the images we used was significantly more attractive than its counterpart.

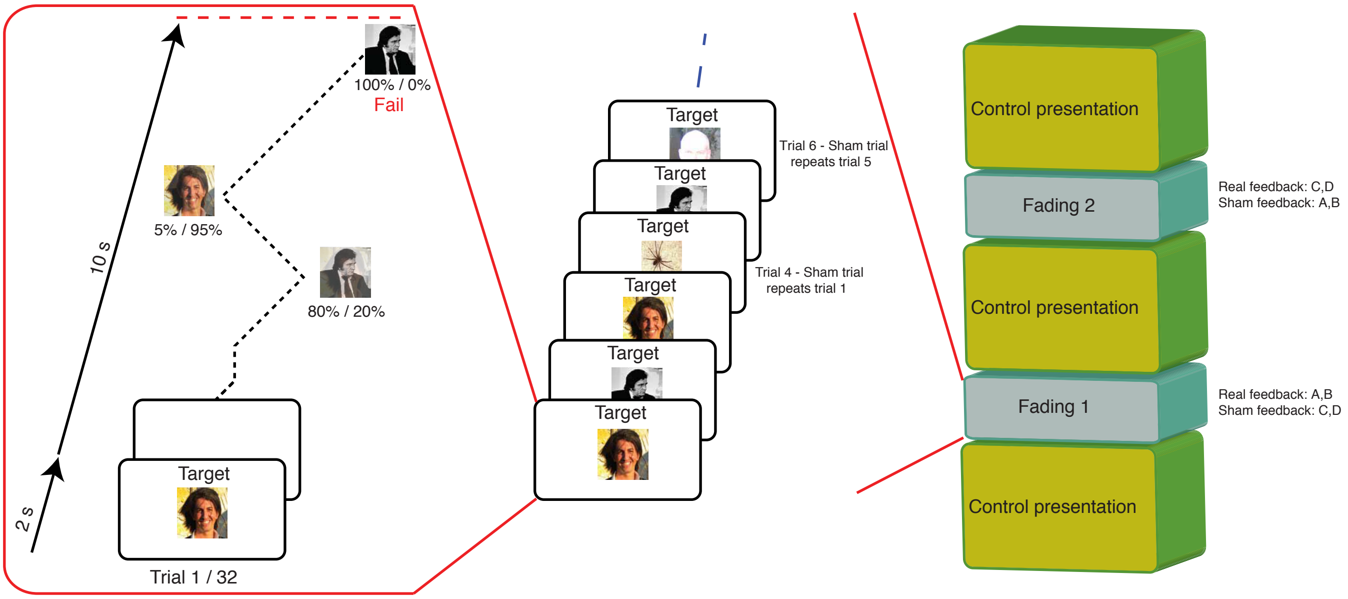


Figure 1. Illustration of the experiment. Right panel shows an illustration of the entire experiment, broken into 5 blocks. The experiment had 3 repetitions of the control presentation (blocks 1, 3, 5), and 2 fading blocks (2 and 4). Real feedback in block 2 was given to two out of the four units, while the remaining two received feedback from a previous trial (sham feedback). The pairs alternated in block 4. **Central panel** shows an illustration of 6 targets in a fading block, corresponding to fading 1 on the right. While in this example the subject is receiving feedback coming directly in real-time from four MTL units in his head that respond selectively to pictures of Johnny Cash and the first author, he receives false feedback (from a previous trial) for the picture of the spider and the man. **Left panel** illustrates a single trial in the experiment (corresponding to a single trial in the central panel), where the subject had the first author as his target, after which he faded in and out of images of the author and Johnny Cash until he reached a 100% visual presentation of Johnny Cash ('failed' trial).

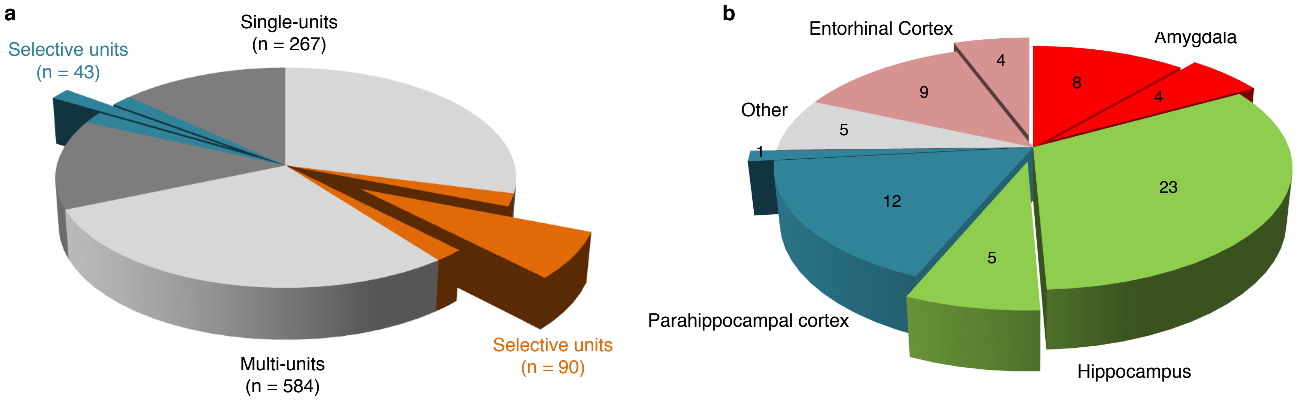


Figure 2. Units distribution. a) Out of 851 units recorded in the course of 18 sessions with 12 subjects, we identified 584 multi-units (69%) and 267 single-units (31%). Out of these, 133 (90 multi-units and 43 single-units) were responsive to one or more images in a prior screening. From these selective units, we used 58 multi-units and 14 single-units for fading. **b)** The 72 units used in the fading experiments, distributed by regions. The exploded slices represent single-units for each region. Regions titled “Other” include: left and right anterior cingulate gyrus, right posterior cingulate gyrus, left temporal occipital and right occipital lobes.

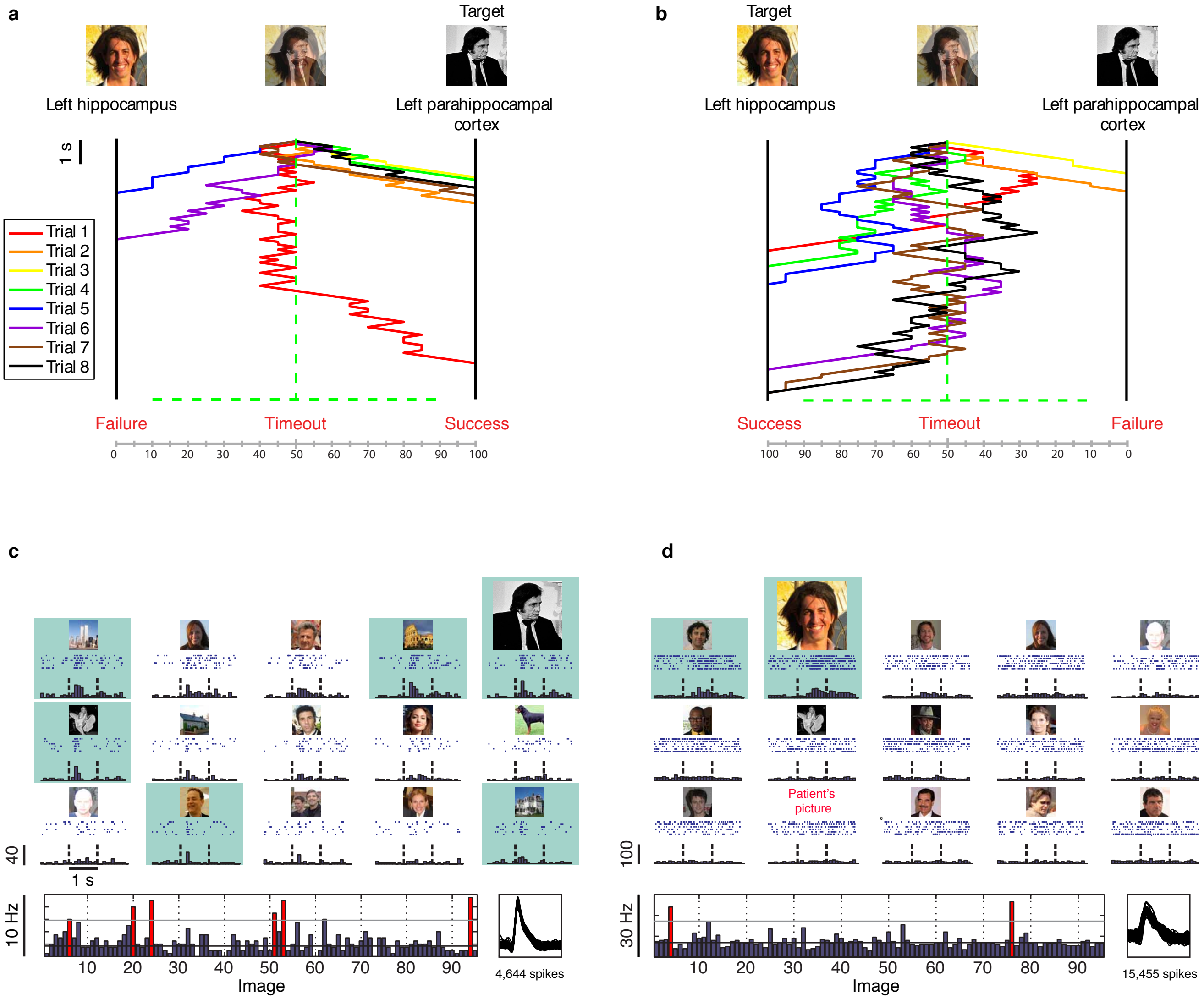


Figure 3. Example of a single block from one subject. Top: Data from 16 trials in which a photo of the first author (with a corresponding neuronal response from a single-unit in the left hippocampus, baseline firing rate: 11.7Hz, firing rate during screening: 25.6Hz, firing rate during fading: 21.0Hz, TDC: 0.69) was superimposed onto a picture of Johnny Cash (with a corresponding neuronal response of a single-unit in the left parahippocampal cortex, baseline firing rate: 4.9Hz, firing rate during screening: 19.4Hz, firing rate during fading: 18.1Hz, TDC: 0.63). Throughout the course of the experiments, the subject had become familiar with the first author. The top-left panel illustrates the 8 trials during which the subject had to make Johnny Cash's images dominate and the first author's image fade away. Time runs downward. Each trial corresponds to a color-coded line of steps that, in turn, correspond to decoding of 100ms firing rates of 4 units. Essentially, the decoder determines whether the activity of the 4 units is close to the activity associated with a photo of the first author, or of Jonny Cash or of neither. The subject was able to fade the image into Johnny Cash's picture 6 out of 8 times. The right panel shows the 8 trials in which the subject was asked to move towards the image of the first author. The subject succeeded in 6 out of 8 trials. No timeout occurred in any of the 16 trials. **Bottom:** Responses to 15 of the 95 images from the units in the left parahippocampal cortex (**left panel**) and left hippocampus (**right panel**) during the screening session. There were no statistically significant responses to the other 80 pictures. For each picture, the corresponding raster plots (six trials are ordered from top to bottom) and post-stimulus time histograms are given. Vertical dashed lines indicate image onset and offset (1s apart). Lower panel shows the mean firing rate during image presentation for all images. The two horizontal lines show the mean baseline activity and the mean plus 5 standard deviations. The corresponding pictures which were deemed responsive are denoted by red bars and highlighted with a grey rectangle. On the right of each panel are the spikes shapes. The spikes histograms in this bottom panel correspond to the sorted spikes, as they correspond to the morning screening session, unlike the upper plot which corresponds to multi-units used in the real-time fading experiment. The two images selected for the following fading experiments are enlarged.

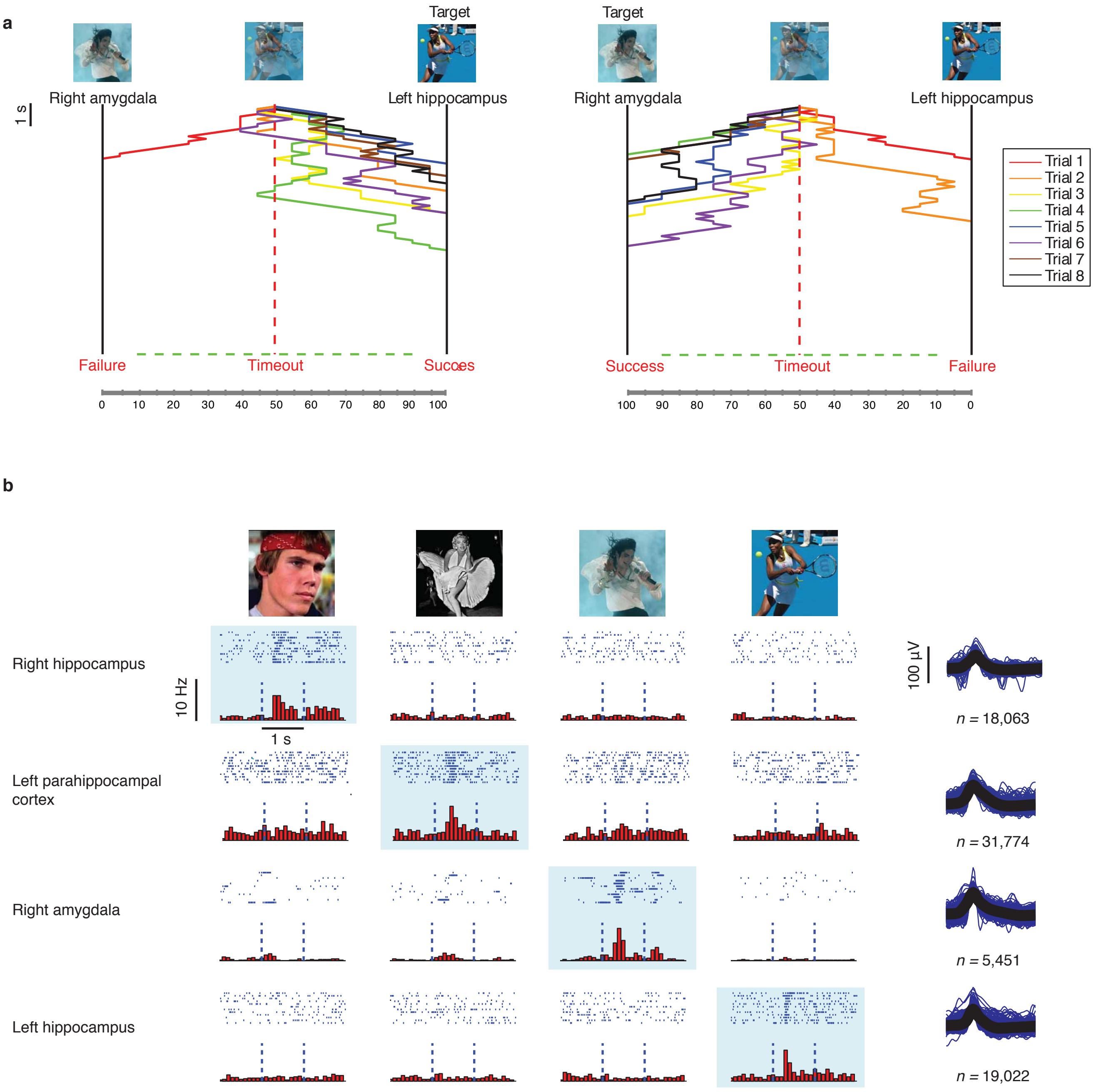


Figure 4. Example of a single session. a) Example of 16 trials for subject 6, where the targets were Michael Jackson (with a responsive single-unit in the subject's right amygdala) and the tennis player Venus Williams (with a responsive multi-unit in the left hippocampus). The subject succeeded in fading in 6 out of 8 trials for Michael Jackson and in 7 out of 8 trials for Venus Williams. No timeout occurred. **b)** The responses in the 4 channels used for the fading experiment, during the preceding control presentation, where each of the images was presented for 12 times. Each channel was exclusively responsive to a single image. On the right are the spike shapes and the number of spikes during the experiment. Channel 15 (right hippocampus, baseline firing rate: 2.2Hz, firing rate during screening: 11Hz, firing rate during fading, when the unit's preferred stimulus is the target: 6.1Hz, Top-Down Control (TDC): 0.63); Channel 53 (left parahippocampal cortex, baseline firing rate: 4.0Hz, firing rate during screening: 18.2Hz, firing rate during fading, when the unit's preferred stimulus is the target: 14.8Hz, TDC: 0.73); Channel 3 (right amygdala, baseline firing rate: 0.5Hz, firing rate during screening: 4.5Hz, firing rate during fading: 2.3Hz, TDC: 0.04); Channel 42 (left hippocampus, baseline firing rate: 2.3Hz, firing rate during screening: 13.4Hz, firing rate during fading: 7.4Hz, TDC: 0.78).

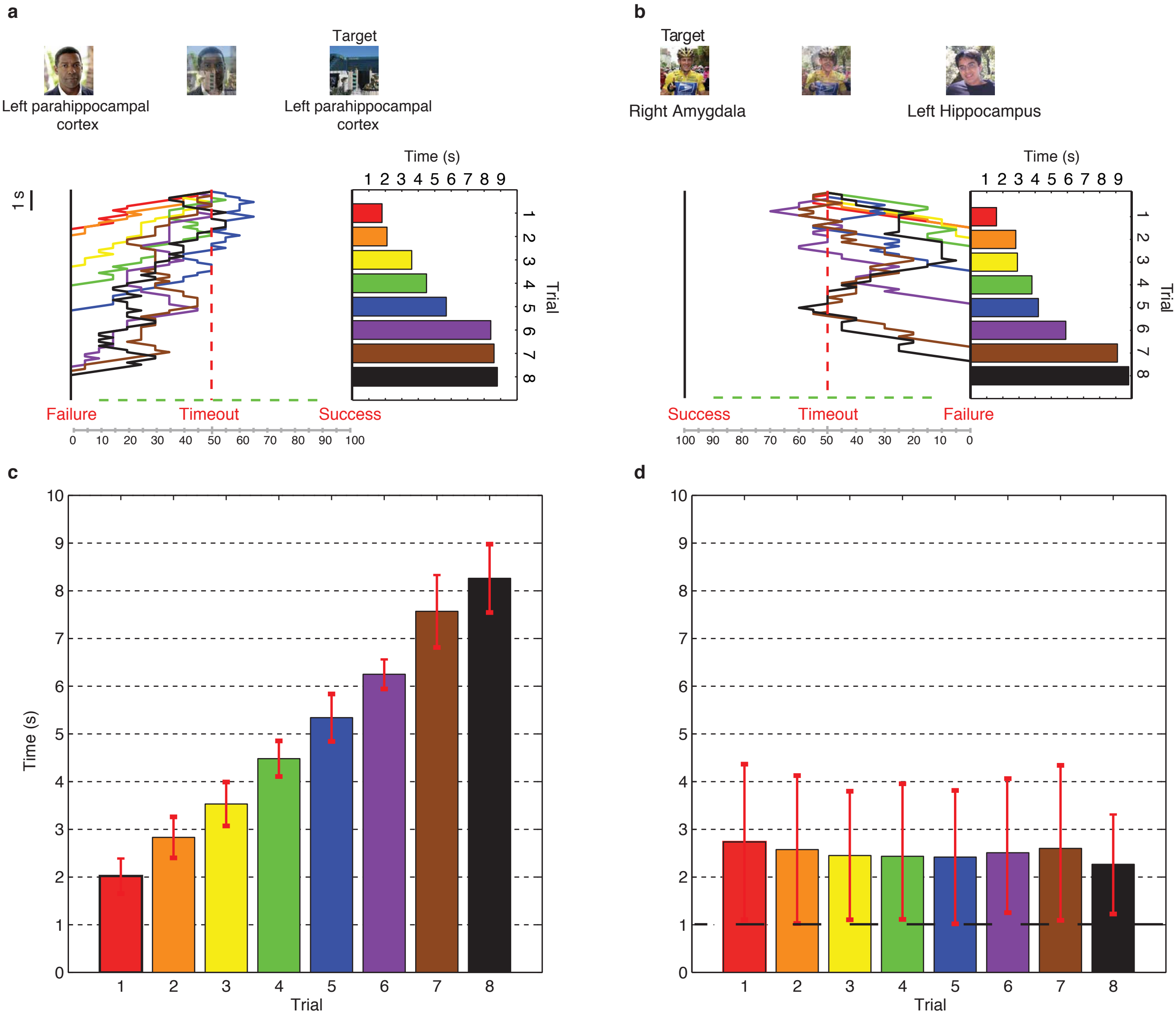


Figure 5. Learning to delay failure. a) Example of learning in a single subject. This subject failed to enhance the image of a building in Las Vegas that the subject was familiar with 8 times. On the right is a corresponding color-coded bar diagram of the time-to-fail - defined as the time to the complete visibility of the distracting image. The subject was able to delay the failure longer on each trial **b)** The fading walk diagram for a different subject who failed in all 8 trials when 'Lance Armstrong' was competing against one of the lab member's pictures. While the first trial failed in a mere 1.60s, failure was delayed for 9.80s in the last trial (timeout occurs after 10s). **c)** Average of the total trial times for all 8 block in 6 subjects with 8 consecutively failed trials. X axis indicates trial number and Y axis the mean time-to-fail for each trial. Red error-bars indicate standard deviation. **d)** Average of the total trial times for 12 blocks in 7 subjects who had 8 consecutive successful trials. While time-to-fail increases significantly, time-to-success remained constant. The thick dashed black line at 1s indicates the minimum trial length possible. The average duration of all 596 successful trials is 2.28 ± 0.85s.

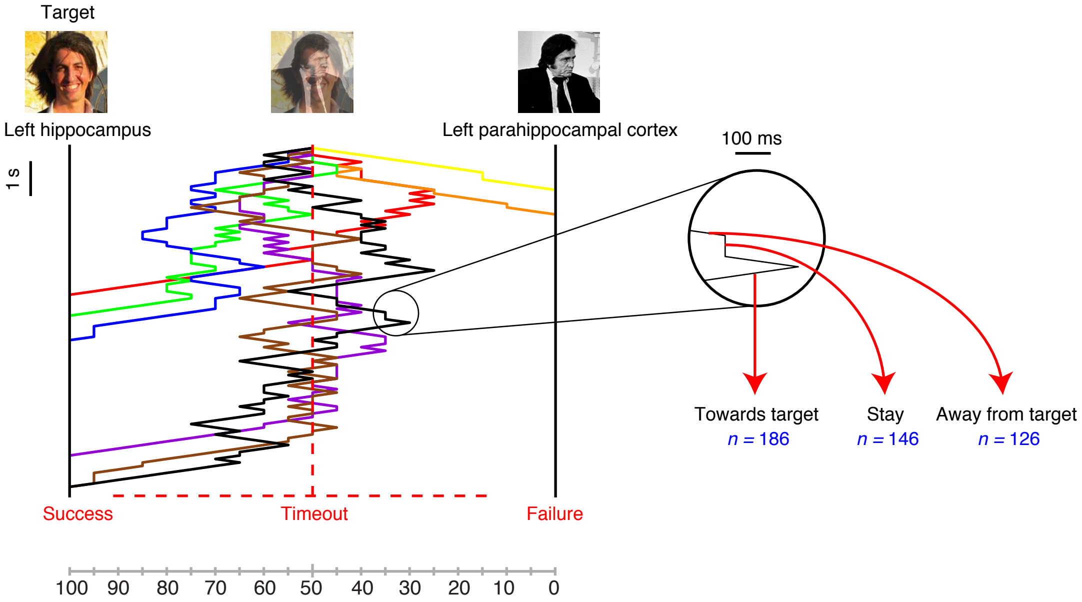


Figure 6. Illustration of the bootstrapping technique. Left panel shows the fading walk for 8 trials in another subject. The firing rate for each unit during fading, divided into 100ms intervals, was classified into one of three categories: 'Towards target' (when the decoding resulted in enhanced visibility of the target), 'Away from target' (visibility of the target decreased) or 'Stay' (no change in visibility). This subject took 186 steps towards the target, 126 steps away, and remained equally far away (stay) during 146 steps, reaching the target in 6 out of 8 trials. We used these proportions as *a priori* probabilities in a Monte-Carlo procedure to create a typical realization of 8 new trials. We repeated this procedure 1000 times and tested how many of the 1000 trials showed lower performance than the observed one.

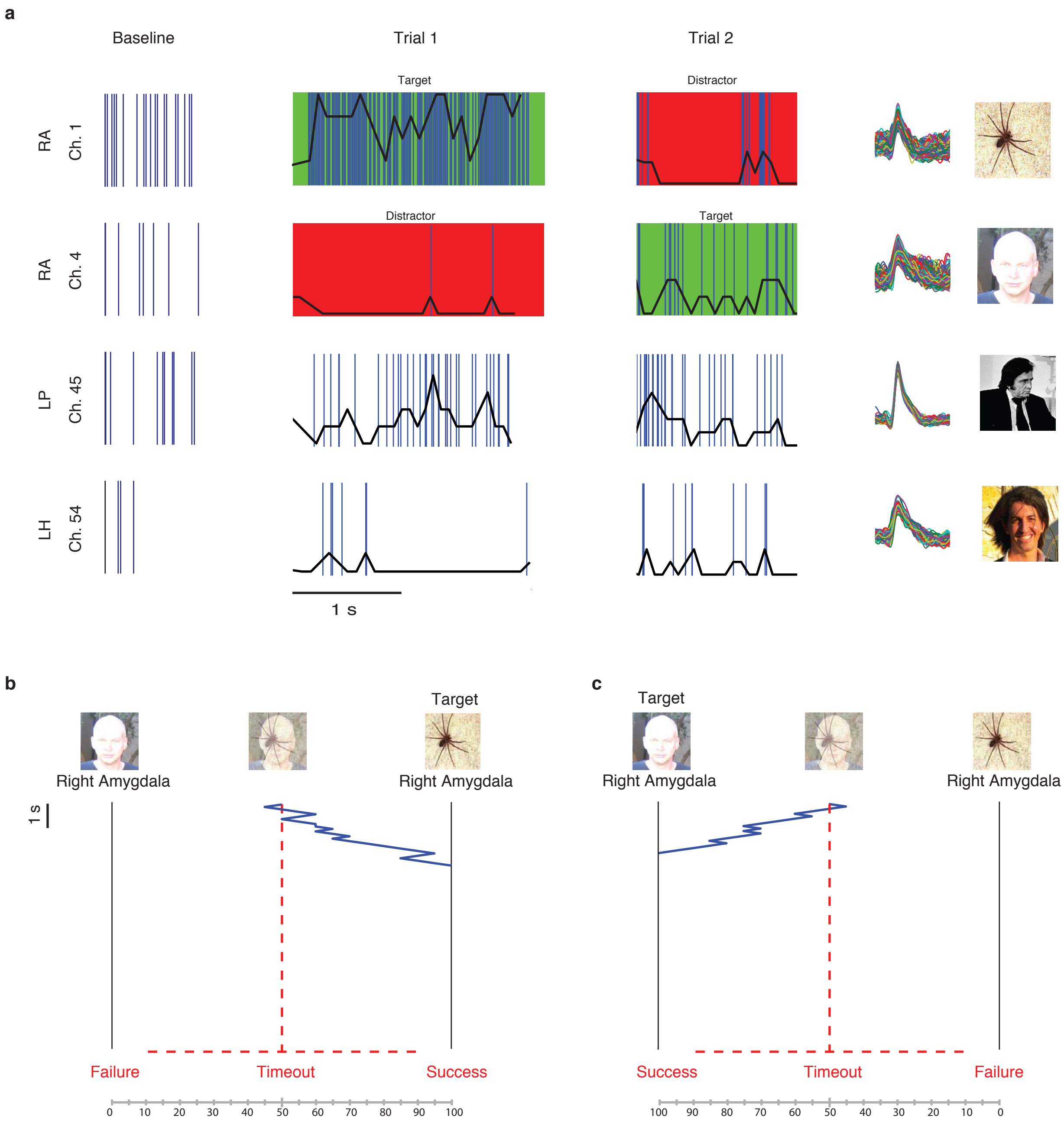


Figure 7. Single trial examples. a) Spiking during the first and the second trials of one subject fading an image of a lab member against an image of a spider. Each row represents an individual unit. Left column illustrates exemplary spikes from a period used for the baseline activity calculation. The second column shows the activity of the 4 units during a trial where the spider was the target. Black line reflects the smoothened spike density function, graphically illustrating the increase above baseline of the unit on channel 1, and decrease below baseline of the unit on channel 4. The third column corresponds to the next trial, where the same pair was pitted against each other but the target was the lab member. The fourth column plots the spike shapes after spike sorting and the last column the preferred image for each unit. **b)** and **c)** illustrate the trial trajectories along the conceptual images plane.

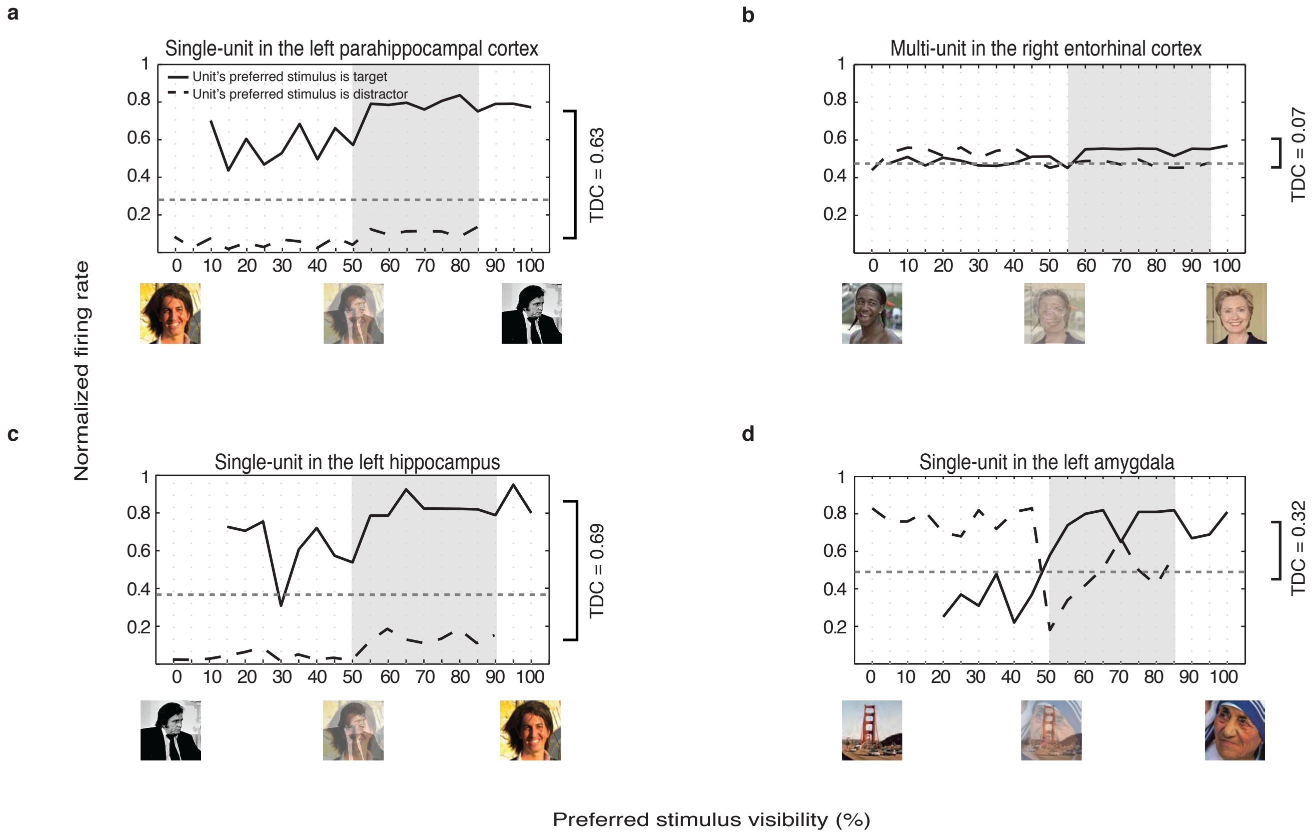
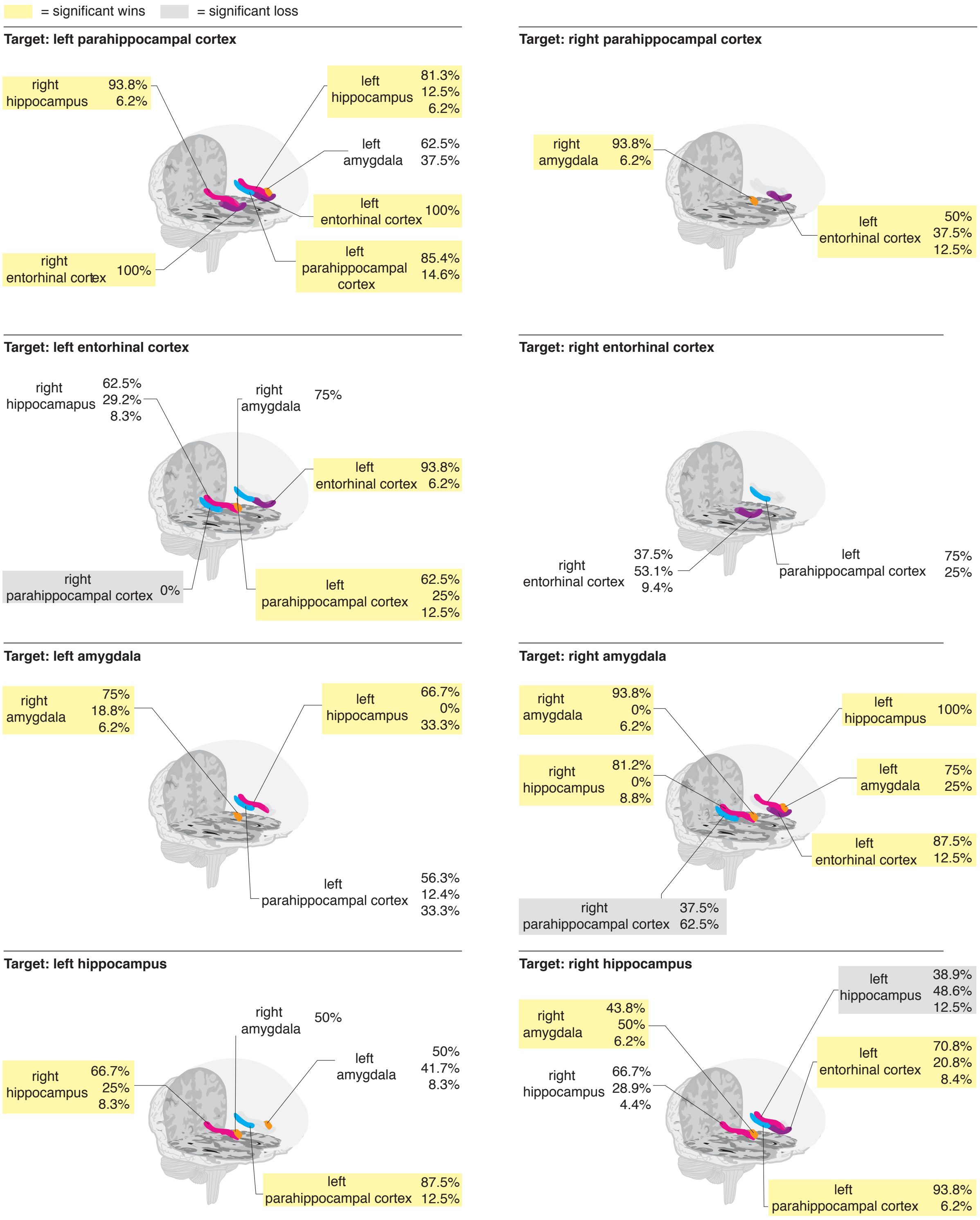


Figure 8. Additional examples of Top-Down-Control. The normalized firing rates of six units in three subjects (panels **a** and **c** are from the subject shown in Supplementary Fig. 3) and sessions as a function of transparency, as in Fig. 4. The top-down-control index for each unit is shown on the right. These curves are typical for these four regions. That is, cognitive control was typically strong in hippocampus and parahippocampal cortex and weak in the amygdala.



= significant wins
= significant loss
Target: left parahippocampal cortex
right hippocampus 93.8% 6.2%
left hippocampus 81.3% 12.5% 6.2%
left amygdala 62.5% 37.5%
left entorhinal cortex 100%
right entorhinal cortex 100%
left parahippocampal cortex 85.4% 14.6%
Target: right parahippocampal cortex
right amygdala 93.8% 6.2%
left entorhinal cortex 50% 37.5% 12.5%
Target: left entorhinal cortex
right hippocamapus 62.5% 29.2% 8.3%
right amygdala 75%
left entorhinal cortex 93.8% 6.2%
right parahippocampal cortex 0%
left parahippocampal cortex 62.5% 25% 12.5%
Target: right entorhinal cortex
right entorhinal cortex 37.5% 53.1% 9.4%
left parahippocampal cortex 75% 25%
Target: left amygdala
right amygdala 75% 18.8% 6.2%
left hippocampus 66.7% 0% 33.3%
left parahippocampal cortex 56.3% 12.4% 33.3%
Target: right amygdala
right amygdala 93.8% 0% 6.2%
left hippocampus 100%
right hippocampus 81.2% 0% 8.8%
left amygdala 75% 25%
left entorhinal cortex 87.5% 12.5%
right parahippocampal cortex 37.5% 62.5%
Target: left hippocampus
right amygdala 50%
left amygdala 50% 41.7% 8.3%
right hippocampus 66.7% 25% 8.3%
left parahippocampal cortex 87.5% 12.5%
Target: right hippocampus
left hippocampus 38.9% 48.6% 12.5%
right amygdala 43.8% 50% 6.2%
left entorhinal cortex 70.8% 20.8% 8.4%
right hippocampus 66.7% 28.9% 4.4%
left parahippocampal cortex 93.8% 6.2%

Figure 9. Competition between units across regions. For each of the four regions used in the analysis, we quantified the proportion of trials in which a given unit within a region wins the competition against a unit from a competing region. 'Target' indicates all those units within a specific region whose preferred stimulus was the target. These trials are collapsed across all subjects, and reflect regional differences. For example (upper left), in all the trials where a unit in the left parahippocampal cortex (LP) was pitted against a unit in the left hippocampus (LH), and the parahippocampal unit's preferred stimulus was the target, 81.3% of all trials were successful, 12.5% were failures and 6.2% resulted in a timeout; this difference between LP and LH was significant (sign-test, $p < 0.01$), marked by a yellow rectangle. Competitions where failed trials were significantly in the majority are marked with grey shaded rectangle.

Supplementary Video

Video 1. An example of a feedback experiment. The movie has three parts. The first part shows the control presentation, first of the multi-unit in the right hippocampus, whose preferred stimulus is Marilyn Monroe, and subsequently of a multi-unit in the left parahippocampal cortex whose preferred stimulus is the actor Josh Brolin. Spikes from the two units generate two distinct-sounding beeps. The two units are preferentially activated during viewing of their preferred stimulus. Following each 1s presentation of each image in the control presentation, the subject had to answer whether or not the picture showed a person (not shown in the movie). Part two shows a sequence of trials from the actual experiment. Each trial starts off with the target image shown for 2s, following by the hybrid image comprising a 50%/50% superposition of the target and the distractor. Each such fading movie is controlled by the relative firing activity of four distinct units. At the end of each trial, the subject heard a sound indicating success or failure/timeout. Part three shows the 16 Monroe-Brolin trials in the order they appeared in the experiment. On the right is the actual visual feedback given to the subject. On the left – the corresponding dynamics in the space spanned by the two images (as in Figure 2).

Supplementary references

1. *Quian Quiroga, R., Z. Nadasdy, and Y. Ben-Shaul, Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural computation, 2004.* ***16****(8): p. 1661-1687.*
2. *Waydo, S., et al., Sparse representation in the human medial temporal lobe. Journal of Neuroscience, 2006.* ***26****(40): p. 10232.*
3. *Itti, L. and C. Koch, Computational modeling of visual attention. Nature Reviews Neuroscience, 2001.* ***2****(3): p. 194-203.*