LETTER

doi:10.1038/nature09510

On-line, voluntary control of human temporal lobe neurons

Moran Cerf[1,2,3], Nikhil Thiruvengadam[1,4], Florian Mormann[1,5], Alexander Kraskov[1], Rodrigo Quian Quiroga[1,6], Christof Koch[1,7]* & Itzhak Fried[2,8,9,10]*

Daily life continually confronts us with an exuberance of external, sensory stimuli competing with a rich stream of internal deliberations, plans and ruminations. The brain must select one or more of these for further processing. How this competition is resolved across multiple sensory and cognitive regions is not known; nor is it clear how internal thoughts and attention regulate this competition[1–4]. Recording from single neurons in patients implanted with intracranial electrodes for clinical reasons[5–9], here we demonstrate that humans can regulate the activity of their neurons in the medial temporal lobe (MTL) to alter the outcome of the contest between external images and their internal representation. Subjects looked at a hybrid superposition of two images representing familiar individuals, landmarks, objects or animals and had to enhance one image at the expense of the other, competing one. Simultaneously, the spiking activity of their MTL neurons in different subregions and hemispheres was decoded in real time to control the content of the hybrid. Subjects reliably regulated, often on the first trial, the firing rate of their neurons, increasing the rate of some while simultaneously decreasing the rate of others. They did so by focusing onto one image, which gradually became clearer on the computer screen in front of their eyes, and thereby overriding sensory input. On the basis of the firing of these MTL neurons, the dynamics of the competition between visual images in the subject's mind was visualized on an external display.

One can direct one's thoughts via external stimuli or internal imagination. Decades of single-neuron electrophysiology and functional brain imaging have revealed the neurophysiology of the visual pathway[1,2]. When images of familiar concepts are present on the retina, neurons in the human MTL encode these in an abstract, modality-independent[5] and invariant manner[6,7]. These neurons are activated when subjects view[6], imagine[8] or recall these concepts or episodes[9]. We are interested here in the extent to which the spiking activity of these neurons can be overridden by internal processes, in particular by object-based selective attention[10–12]. Unlike imagery, in which a subject imagines a single concept with closed eyes, we designed a competitive situation in which the subject attends to one of two visible superimposed images of familiar objects or individuals. In this situation, neurons representing the two superimposed pictures vie for dominance. By providing real-time feedback of the activity of these MTL neurons on an external display, we demonstrate that subjects control the firing activity of their neurons on single trials specifically and speedily. Our subjects thus use a brain–machine interface as a means of demonstrating attentional modulation in the MTL.

Twelve patients with pharmacologically intractable epilepsy who were implanted with intracranial electrodes to localize the seizure focus for possible surgical resection[13] participated. Subjects were instructed to play a game in which they controlled the display of two superimposed images via the firing activity of four MTL units in their brain (Fig. 1). In a prior screening session, in which we recorded activity from MTL regions that included the amygdala, entorhinal cortex, parahippocampal cortex and hippocampus, we identified four different units that responded selectively to four different images[6]. Each trial started with a 2-s display of one of these four images (the target). Subjects next saw an overlaid hybrid image consisting of the target and one of the three remaining images (the distractor), and were told to enhance the target ('fade in') by focusing their thoughts on it. The initial visibility of both was 50% and was adjusted every 100 ms by feeding the firing rates of four MTL neurons into a real-time decoder[14] that could change the visibility ratios until either the target was fully visible ('success'), the distractor was fully visible ('failure'), or until 10 s had passed ('timeout'; see Fig. 2, Supplementary Figs 3 and 4 and Supplementary Video). We considered subjects' 'trajectories' in the plane defined by time and by the transparency of the two images making up the hybrid (Fig. 2a).

The subjects manipulated the visibility of the hybrid image by any cognitive strategy of their choosing. Six out of 12 subjects reported in a follow-up interview that they focused on the concept represented by the target picture (most often a person) or closely allied associations. Subjects did not employ explicit motor strategies to control these four units (see Supplementary Information). Subjects participated without any prior training and with a striking success rate in a single session lasting around 30 min, reaching the target in 596 out of 864 trials (69.0%; 202 failures and 66 timeouts). Results were significant ($P < 0.001$, Wilcoxon rank-sum) for each subject (Fig. 3). Subjects successfully moved from the initial 50%/50% hybrid image to the target in their first trial in 59 out of 108 first trials (54.6%).

Testing the extent to which successful competition between the two units responsive to the two images depends on their being located in different hemispheres, in different regions within the same hemisphere or within the same region (Fig. 3b), revealed that 347 out of 496 trials involving inter-hemispheric competitions were successful (70.0%; 123 failures, 26 timeouts), 177 out of 256 intra-hemispheric but inter-regional competitions were successful (69.1%; 45 failures, 34 timeouts) and 72 out of 112 intra-regional competitions were successful (64.0%; 30 failures, 10 timeouts). There is no significant difference between these groups at the $P = 0.05$ level.

Every 'fading sequence' in each trial that every subject saw was based entirely on the spiking activity of a handful of neurons in the subject's brain. We recorded from a total of 851 units, of which 72 were visually responsive (see ref. 6 for definition of 'responsive') and were used for feedback. In light of the explicit cognitive strategies reported by subjects—enhancing the target and/or suppressing the distractor—the question arises whether successful fading was due to increasing firing

[1]Computation and Neural Systems, California Institute of Technology, Pasadena, California 91125, USA. [2]Department of Neurosurgery, University of California, Los Angeles, California 90095, USA. [3]Stern School of Business, New York University, New York, New York 10012, USA. [4]School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA. [5]Department of Epileptology, University of Bonn, Bonn 53105, Germany. [6]Department of Engineering, University of Leicester, Leicester LE1 7RH, UK. [7]Department of Brain and Cognitive Engineering, Korea University, Seoul, 136-713, Korea. [8]Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California 90095, USA. [9]Functional Neurosurgery Unit, Tel-Aviv Medical Center, Tel-Aviv 64239, Israel. [10]Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv 69978, Israel.

*These authors contributed equally to this work.

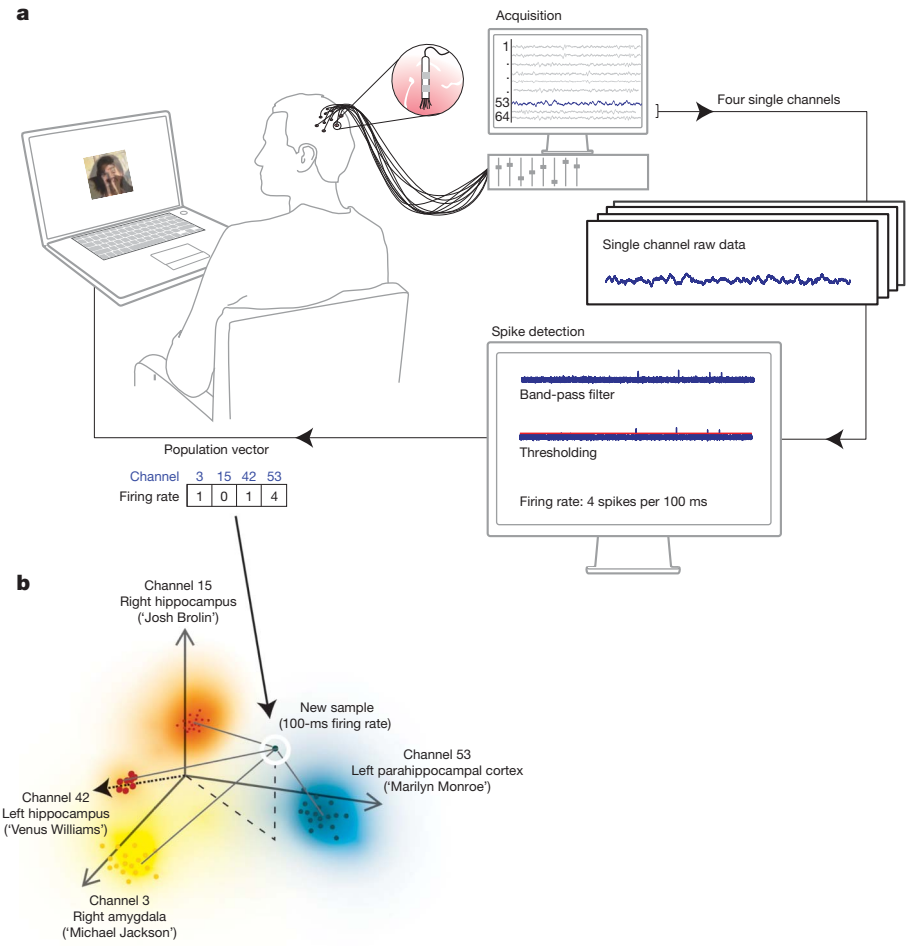

Figure 1 | Experimental set-up. a, Continuous voltage traces are recorded by 64 microelectrodes from the subject's medial temporal lobe. A four-dimensional vector, corresponding to the number of action potentials of four responsive units in the previous 100 ms, is sent to a decoding algorithm determining the composition of the hybrid seen by the subject with a total delay of less than 100 ms. **b**, The closest distance (weighted by the standard deviation) of this vector to the four clusters representing the four images is computed. If the 'winning' cluster represents the target or the distractor image, the visibility ratio of these two is adjusted accordingly.

of the unit the preferred stimulus of which was the target, to reducing the activity of the unit the preferred stimulus of which was the distractor or a combination of both. To answer this, we calculated firing rates in 100-ms bins in each trial for each unit. These rates were assigned to one of three categories labelled as follows. 'Towards target' meant the decoding process (based on the firing rate of all four units in this bin) enhanced the visibility of the target image, 'Away from target' meant decoding enhanced the distractor image and 'Stay' meant no change in visibility occurred (Supplementary Fig. 6). In the majority of successful trials (84.6%), the firing rate of the target-preferring unit was enhanced (3.72 standard deviations above baseline, $P < 10^{-4}$, *t*-test; Supplementary Fig. 7), simultaneously with suppression of the distractor-preferring unit (0.59 standard deviations below baseline, $P < 10^{-4}$, *t*-test). In 12.9% of successful trials only enhancement was seen, and in 1.1% only a reduction was seen. In the remaining trials, no significant deviation in baseline was detected. We observed no change in firing rates of the two units used for decoding, whose preferred stimuli were not part of the fading trial. Thus, successful fading was not caused by a generalized change in excitation or inhibition but by a targeted increase and decrease in the firing of specific populations of neurons. No long-lasting effect of feedback on the excitability of the MTL neurons was seen (see Supplementary Information).

To disentangle the effect of the retinal input from the instruction, we compared the activity of each unit in successful trials when the target was the unit's preferred stimulus (target trials) with activity in successful trials when the target was the unit's non-preferred stimulus (distractor trials). This comparison was always done for the same retinal input, measured by the percentage of the visual hybrid allotted to the target (Fig. 4). We normalized each unit's response by its maximal firing rate over the entire experiment, and averaged over all trials for all subjects. For the same retinal input, the firing rate of neurons responding to the target pictures was much higher when subjects focused their attention on the target than when they focused on the distractor. The only difference was the mental state of the subject, following the instruction to suppress one or the other image.

To quantify the extent to which attention and other volitional processes dominate firing rates in the face of bottom-up sensory evoked responses, we devised a top-down control (TDC) index. TDC quantifies the level of control that subjects have over a specific unit and is the difference between the normalized firing rate when the subject attended the unit's preferred stimulus and the normalized rate when the subject attended the distractor image. That is, we subtracted the lower from the upper curve in Fig. 4a. Averaged over all 72 units, TDC equals 0.44 ± 0.28 (mean $\pm$ standard deviation), highly significantly different from zero. This was not true for failed trials (mean $P = 0.18$). If instead of subtracting the two curves the upper curve is divided by the lower one, a ratio of 6.17 ± 5.02 is obtained, highly significantly different from one. That is, the average unit fires more than six times as

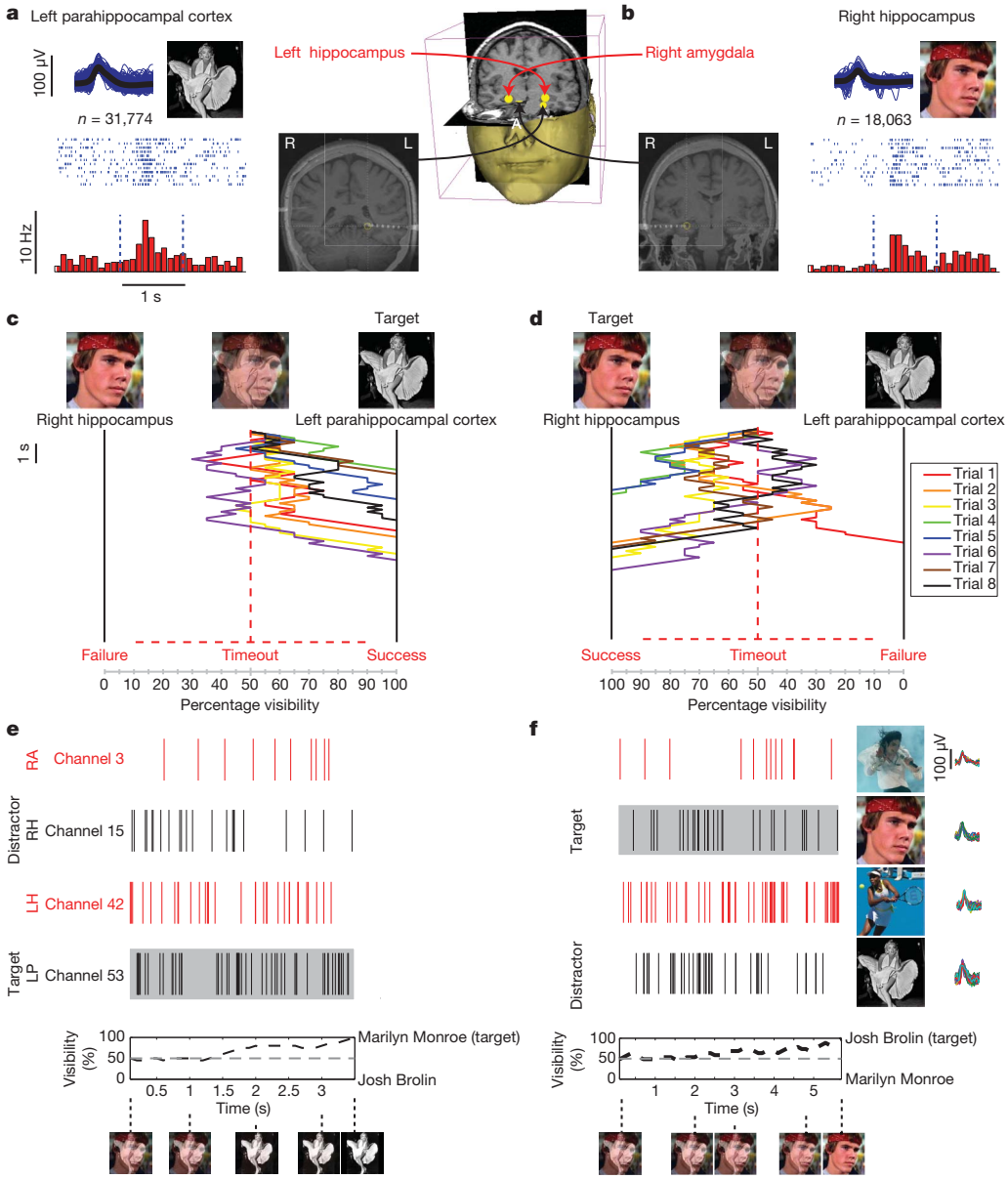

Figure 2 | Task performance and neuronal spiking. Two American actors, 'Josh Brolin' and 'Marilyn Monroe', constituted the preferred stimulus for two units. **a**, One multi-unit responded selectively to Monroe and was located in the left parahippocampal cortex. Below each illustration are the corresponding raster plots (twelve trials are ordered from top to bottom) and post-stimulus time histograms obtained during the control presentation. Vertical dashed lines indicate image onset (left) and offset (right), 1-s apart. Spike shapes are shown in blue, and the average spike shape in black. Below are the total number of spikes during the session. On the right is an illustration of the brain regions competing in these trials, and a fusion of the coronal CT and MRI scans taken after electrode implantation. Here, competing units were located in different hemispheres and regions. See Supplementary Video of the actual experiment. **c**, Time (running downwards for 10 s) versus percentage visibility of eight trials in which the subject had to fade a 50%/50% hybrid image into a pure Monroe image. The subject was able to do so all eight times, even though these were her first trials ever. **b**, **d**, When Brolin was the target, she succeeded seven out of eight times. All subjects show similar trends of controlled fading (Fig. 3). The hybrid image was controlled in real time by the spiking of four units selective to the image of Brolin, Monroe, Michael Jackson or Venus Williams. **e**, **f**, Spiking activity of all four units for one successful Monroe (**e**) and Brolin (**f**) trial. The spike shapes and the four images each unit is selective to are shown on the right. Below are the images as seen by the subject during the trial at different times. For another example, see Supplementary Figs 4 and 7. For copyright reasons, some of the original images were replaced in this and all subsequent figures by very similar ones (same subject, similar pose, similar colour and so on). The image of Josh Brolin is copyright The Goonies, Warner Bros. Inc. RA, right amygdala; RH, right hippocampus; LH, left hippocampus; LP, left parahippocampal cortex.

vigorously when the subject is attending to the unit's preferred image than when he/she is attending to the distractor. Excitation of the target unit, alongside inhibition of the distractor unit, occurs even in trials where the distractor is dominating the hybrid image, suggesting that the units are driven by voluntary cognitive processes capable of overriding distracting sensory input.

To control the extent to which successful 'fading in' was caused by the overall level of effort and attentional focus of the subject or by the instantaneous firing activity of the four units, we compared performance during normal feedback to that reached during sham feedback, when the image's visibility was, in fact, not guided by the subject's immediate neuronal activity but by activity from a previous trial (see Methods). Although subjects' level of effort and attention were the same as during real feedback, success dropped precipitously from 69.0% to 31.2% (33.7% failures and 35.1% timeouts; $\chi^2 = 69.9$, degrees of freedom = 2, $P < 10^{-4}$). Only two out of 12 subjects did better than chance during

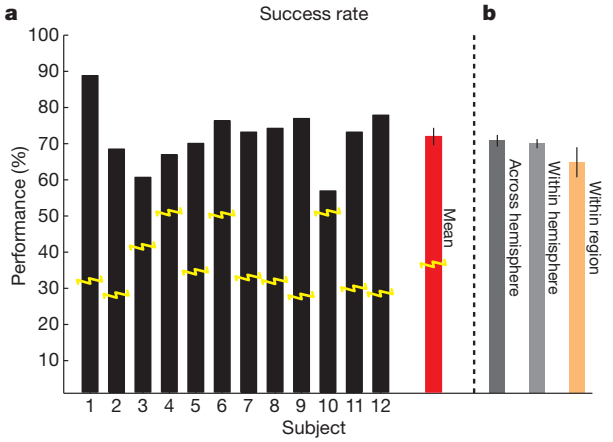

Figure 3 | Successful fading. **a**, Percentage of trials in which subjects successfully controlled the activity of four units and faded to the target image within 10 s. Yellow lines indicate chance performance—determined by bootstrapping 1,000 random trials for each subject ($P < 0.001$; Wilcoxon rank-sum). The red bar is the performance averaged over all 12 subjects. Error bars show the standard deviation. **b**, Percentage of successful trials of the entire data set in which the competition between the two units was across hemispheres, within the same hemisphere but in different regions, or within the same region. Error bars show standard deviations. Note that in **a**, performance is analysed across subjects, whereas in **b** it is analysed across eight trial fading sessions; hence, the means differ.

sham feedback ($P < 0.001$); the rest were not significant (P values: 0.15 ± 0.14). Furthermore, in contrast to the pattern observed with real feedback where subjects were able to successively delay failure over time (Supplementary Fig. 5), there was no such delay during sham feedback (see Supplementary Information). These findings support the notion that feedback from the four selective units controlling the composite image were essential to carry out the task successfully, rather than the general cognitive efforts of the subject, exposure to the stimuli, or global changes in firing activity.

Our study creates a unique design within which to interrogate the mind's ability to influence the dominance of one of two stimuli by decoding the firing activity of four units deep inside the brain. The stronger the activity of the target-preferring unit and the weaker the activity of the distractor-preferring unit, relative to the two other units, the more visible the target became on the screen and the more opaque the superimposed distractor image became (and vice versa). Overall, subjects successfully 'faded-in' 69% of all trials. Cognitive processes voluntarily initiated by the subject, such as focusing on the target or suppressing the distractor, affected the firing activity of four units in different MTL regions, sometimes even across hemispheres (see Supplementary Information for list of all regions). The firing rate of these units generates a trajectory in a four-dimensional space. This was projected onto a one-dimensional walk along a line given by the competing representation of the target and the distractor image and visualized onto an external display. This path that subjects take may be analogous to the movement of rodents navigating in their physical environment using place fields[13].

The past decade has seen major strides in the development of brain–machine interfaces using single-neuron activity in the motor and parietal cortex of monkeys[15–18] and humans[19–22]. A unique aspect of the present study is the provision of feedback from regions traditionally linked to declarative memory processes. It is likely that the rapidity and specificity of feedback control of our subjects depends on explicit cognitive strategies directly matched to the capacity of these MTL neurons to represent abstract concepts in a highly specific yet invariant and explicit manner[5]. We previously estimated, using Bayesian reasoning, that any one specific concept is represented by up to one million MTL neurons, but probably by much less[23]. As our electrodes are sampling a handful of MTL neurons with predetermined selectivities[14], cognitive control

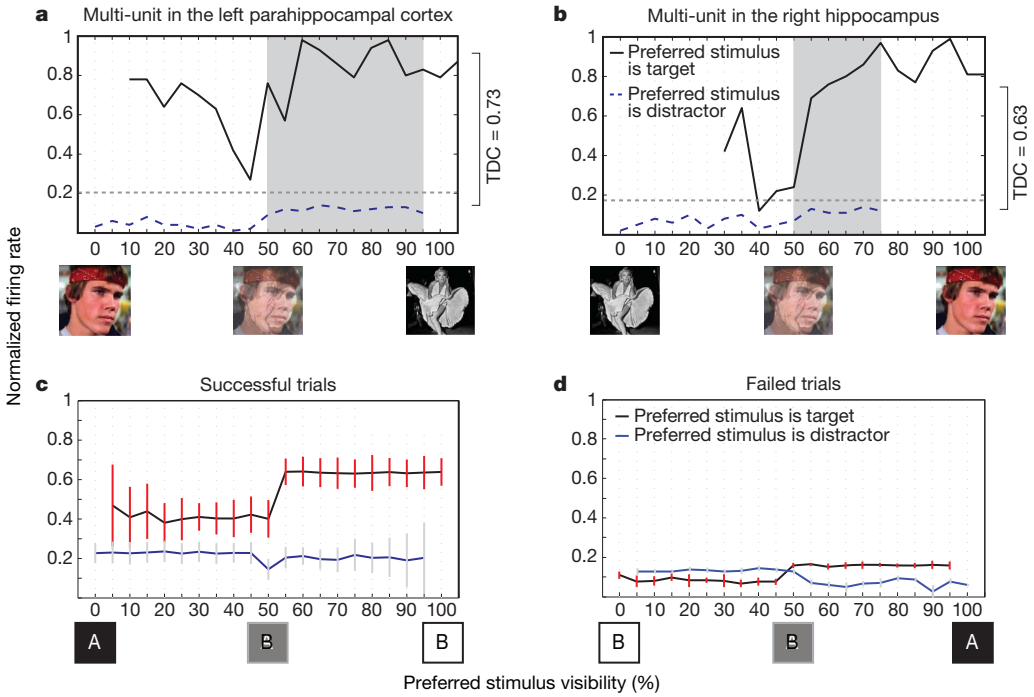

Figure 4 | Voluntary control at the single unit level. **a**, **b**, Normalized firing rates of the units in Fig. 2 as a function of visibility. We averaged the firing rates every 100 ms for every level of visibility for all successful trials where the target either was the unit's preferred (solid, black) or non-preferred stimulus (dashed, blue). Units fired significantly above baseline (grey dashed line) when the target was the preferred stimulus, and less than baseline when the target was the non-preferred stimulus. The TDC index is shown on the right. The shaded area reflects the bins used to calculate TDC. **c**, **d**, Averaging target and distractor trials across all subjects and all units for all successful fading trials reveals that the firing rate is significantly higher when the target is the preferred stimulus than in the competing situation, no matter what the visual input is. This is not true for failed trials (right). Red and dark grey vertical error bars are standard deviations. See Supplementary Fig. 8 for additional examples.

strategies such as object-based selective attention permit subjects to voluntarily, rapidly, and differentially up- and downregulate the firing activities of distinct groups of spatially interdigitated neurons to override competing retinal input. At least in the MTL, thought can override the reality of the sensory input. Our method offers a substrate for a high-level brain–machine interface using conscious thought processes.

METHODS SUMMARY

Subjects. Twelve patients with intractable epilepsy were implanted with depth electrodes to localize the epileptic focus for possible subsequent resection. The placement of all electrodes was determined exclusively by clinical criteria. All patients provided informed consent. All studies conformed to the guidelines of the Institutional Review Boards at UCLA and at Caltech.

Electrophysiology. Extracellular neural activity was acquired using 64 microwires implanted in various regions including the hippocampus, amygdala, parahippocampal cortex, and entorhinal cortex. Selected channels were band-pass filtered at 300–3,000 Hz, and a threshold was applied to detect spikes.

Experimental procedure. In a screening session, approximately 110 images of familiar persons, landmark buildings, animals, and objects were presented six times in random order for 1 s each. Four units were identified, each of which responded selectively to one of four different images. These four images were each presented 12 times to train a decoder. In a following fading experiment, each trial began with a 2-s presentation of the target. The subject then viewed a superposition of the target and one of the remaining three images, and was instructed to "continuously think of the concept represented by that image". Spike counts in 100-ms bins in the four selective units fully controlled the superposition on the screen in real time. At the end of the trial, acoustic feedback was given to the subject indicating success, failure or timeout after 10 s.

Data analysis. To evaluate each subject's performance, we used a bootstrapping technique—generating 1,000 random trials for each set of four units on the basis of their spiking activity and comparing their mean performance to that of the subject. Additionally, we analysed the activity of single and multi-units, compared against sham trials, compared unit activity across different regions, tested for changes in neuronal characteristics over time, and tested the level of control that subjects can exert over their neurons.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 January; accepted 14 September 2010.

1. Chalupa, L., Werner, J. & Barnstable, C. *The Visual Neurosciences* (MIT Press, 2004).
2. Thorpe, S. Single units and sensation: still just as relevant today. *Perception* **38,** 804–807 (2009).
3. Blake, D. A. R. (ed.) *Binocular Rivalry* (MIT Press, 2005).
4. Reynolds, J. & Chelazzi, L. Attentional modulation of visual processing. *Annu. Rev. Neurosci.* **27,** 611–648 (2004).
5. Quian Quiroga, R. *et al.* Explicit encoding of multimodal percepts by single neurons in the human brain. *Curr. Biol.* **19,** 1308–1313 (2009).
6. Quian Quiroga, R. *et al.* Invariant visual representation by single neurons in the human brain. *Nature* **435,** 1102–1107 (2005).
7. Földiák, P. Neural coding: non-local but explicit and conceptual. *Curr. Biol.* **19,** R904–R906 (2009).
8. Kreiman, G., Koch, C. & Fried, I. Imagery neurons in the human brain. *Nature* **408,** 357–361 (2000).
9. Gelbard-Sagiv, H. *et al.* Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* **322,** 96–101 (2008).
10. Reddy, L., Kanwisher, N. & VanRullen, R. Attention and biased competition in multi-voxel object representations. *Proc. Natl Acad. Sci. USA* **106,** 21447–21452 (2009).
11. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18,** 193–222 (1995).
12. Serences, J. *et al.* Control of object-based attention in human cortex. *Cereb. Cortex* **14,** 1346–1357 (2004).
13. Fried, I., MacDonald, K. & Wilson, C. Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron* **18,** 753–765 (1997).
14. Quian Quiroga, R. *et al.* Decoding visual inputs from multiple neurons in the human temporal lobe. *J. Neurophysiol.* **98,** 1997–2007 (2007).
15. Musallam, S. *et al.* Cognitive control signals for neural prosthetics. *Science* **305,** 258–262 (2004).
16. Wessberg, J. *et al.* Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* **408,** 361–365 (2000).
17. Velliste, M. *et al.* Cortical control of a prosthetic arm for self-feeding. *Nature* **453,** 1098–1101 (2008).
18. Moritz, C., Perlmutter, S. & Fetz, E. Direct control of paralysed muscles by cortical neurons. *Nature* **456,** 639–642 (2008).
19. Hochberg, L. *et al.* Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* **442,** 164–171 (2006).
20. Kim, S. *et al.* Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. *J. Neural Eng.* **5,** 455–476 (2008).
21. Kennedy, P. *et al.* Direct control of a computer from the human central nervous system. *IEEE Trans. Rehabil. Eng.* **8,** 198–202 (2000).
22. Guenther, F. *et al.* A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE* **4,** e8218 (2009).
23. Waydo, S. *et al.* Sparse representation in the human medial temporal lobe. *J. Neurosci.* **26,** 10232–10234 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the patients for their participation in these studies. We thank K. Laird, A. Postolova, N. Parikshak and V. Isiaka for help with the recordings; E. Behnke and T. Fields for technical support; G. Mulliken and U. Rutishauser for comments on the manuscript; and M. Moon for help with data visualization. This work was supported by grants from the National Institute of Neurological Disorders and Stroke (NINDS), the National Institute of Mental Health (NIMH), the G. Harold & Leila Y. Mathers Charitable Foundation, and the WCU programme through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-2008-000-10008-0).

Author Contributions M.C., F.M., R.Q.Q., C.K. and I.F. designed the experiment; M.C. performed the experiments; I.F. performed the surgeries; M.C. and N.T. analysed the data; M.C., C.K. and I.F. wrote the manuscript. All authors discussed the data and the analysis methods and contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.C. (moran@klab.caltech.edu), C.K. (koch@klab.caltech.edu) or I.F. (ifried@mednet.ucla.edu).

METHODS

Subjects. Twelve patients participated in the study. Patients had pharmacologically intractable epilepsy and had been implanted with depth electrodes to localize the epileptic focus for possible subsequent resection. For each patient, the placement of the depth electrodes, in combination with microwires, was determined exclusively by clinical criteria[13]. All patients provided informed consent. All studies conformed to the guidelines of the Medical and Human subjects Institutional Review Boards at UCLA and the California Institute of Technology.

Screening. An initial morning screening session was recorded, during which approximately 110 images of familiar persons, landmark buildings, animals, and objects were presented six times in random order for 1 s each, after which each subject was asked to indicate with a button press whether the image contained a person or not. A standard set of such images was complemented by images chosen after an interview with the subject that determined which celebrities, landmarks, animals and objects the subject might be most familiar with. This approximately 30-min-long session—110 images × 6 repetitions × (1 s + reaction time)—was evaluated off-line to determine which of the 110 images elicited a response in at least one of 64 recorded channels, based on the criteria outlined in ref. 6. This involves measuring the median firing rate during the 300–1,000 ms after image onset across the six repetitions and comparing it to the baseline activity of the channel from 1,000–300 ms before image onset. Stimuli with median firing rates five standard deviations above baseline were considered selective.

From the group of selective units we chose four, based on their selectivity. The general guidelines for selection were: (1) to choose units from different brain regions so as to allow for competition between regions, (2) to select units that had similar characteristics in terms of latency and duration of the response within the 1 s the selective image is onscreen, and (3) to choose units for which the difference between firing rate during presentation and baseline was particularly clear. This selection was done by eye and was not quantitative.

Control presentations. The fading paradigm began with a short control presentations session—a presentation of the four selected images in random order, 12 repetitions at 1 s each—in a manner exactly replicating the set-up of the earlier screening session (see Supplementary Fig. 4 for results of the first control presentation for four units of one subject). The median firing activity over these 48 presentations between 1,000–300 ms before image onset determined the baseline firing rate for that unit for further statistical comparisons. The data from the control presentation procedure allowed for the set-up of a population-vector-based decoder.

We repeated the control presentation twice during each experiment—between the feedback blocks and at the end of the experiment, to verify that the neurons were still responsive for the stimuli used (Supplementary Fig. 1).

Fading. The following main fading experiment consisted of blocks of 32 trials each: eight for each of the four stimuli, shown in random order. Each trial began with a 2-s presentation of the target image. Subsequently, the subject viewed a superposition of the target image and one of the remaining three images (these two images were paired for the entire block). The hybrid image (H) was constructed from the target (T) and distractor image (D) by:

$$H = \alpha T + (1 - \alpha)D$$

where $\alpha \in [0, 1]$ corresponds to the trajectory in the images space—starting at 0.5 and changing in steps of 0.05 every 100 ms, ending either at 0 or 1 (see Supplementary Fig. 1 for illustration). α was controlled by the decoder, that is, ultimately by four units in the subject's brain.

The subject was instructed to enhance the target image from the hybrid image on the screen by "continuously thinking of the concept represented by that image". The subject was not directed in any further manner on what cognitive strategy to use—such as imagining that particular image or focusing on an aspect of the image—but was encouraged to explore the vast area of thoughts which might elicit a response. At the end of the trial, acoustic feedback was given to the subject indicating success, failure or timeout. The latter occurred after 10 s.

In each fading block (32 trials), two of the four images (say, A and B, together having 16 trials—eight trials with A as the target and eight with B as the target) received sham feedback, which did not reflect the neuronal activity during that trial. There was no overt difference between true and sham feedback trials. To achieve balanced exposure, any sham trial was a direct repetition of one prior real trial. For example, for a sham trial where image A was the target, the subject saw a hybrid image of A and B but the course of changes in each image's visibility was in fact based on the neuronal activity of a different previous trial (say, a trial with image C as the target and D as the distractor).

Decoding. Data from four selected channels (microwires) were read, and spikes were detected in real time for every 100-ms interval during the control presentation. Each 1-s image presentation in the control presentation (four images × 12 repetitions) was broken into ten 100-ms bins. We used spikes from the seven bins from 300 ms to 1,000 ms following image onset for the analysis because these included the most relevant data for decoding[14]. The total numbers of spikes for each 100-ms bin formed clusters in a four-dimensional space representing the activity of the four units for each image. Thus, for 12 (repetitions) × 4 (images) × 7 (bins) we obtained a 336 (cluster) by 4 (channels) matrix corresponding to the firing rate during each image presentation for all 100-ms bins.

During fading, the firing rates from the four channels gave rise to a population vector that was used to associate the corresponding 100-ms bin to one of the four images. The population vector was a point in four-dimensional space, and we used the Mahalanobis distance to determine which cluster the point was closest to. The Mahalanobis distance was chosen as the distance measure because it is a fast and linear distance calculation measure that takes into account the shape of the cluster. Previous data showed that cluster variability is significant for our data[14], so taking the standard deviation of the cluster into account yielded better decoding.

The distance D from each of the four clusters is calculated as:

$$D = (\mathbf{x} - \bar{S}) \times \mathrm{COV}(S)^{-1} \times (\mathbf{x} - \bar{S})^{\mathrm{T}}$$

where $x = (x_1, x_2, x_3, x_4)$ is the new point in the four-dimensional space (corresponding to the firing rate of four units in the previous 100 ms). S is a 336×4 matrix of firing rates of four units during 100-ms bins in the control presentation when the subject was viewing one of four images (for example, columns 1:7 in the matrix correspond to seven 100-ms bins of the firing rates of the four channels while image A was on the screen, columns 8:14 correspond to activity while image C was on the screen, and so on) and $\bar{S}$ is the mean of S. $D = (d_1, d_2, d_3, d_4)$ where d_i corresponds to the distance from cluster i. COV is the covariance function.

The closest cluster was regarded as the concept the subject thought of. Notice that each trial consists of two concepts that, when decoded, directly influenced the visibility of the two associated images that make up the hybrid (annotated as A and B). Decoding of one of the other two concepts (annotated C and D) was interpreted as 'thinking of neither A nor B'. In any given 100 ms of each fading trial, there were three possible outcomes: (1) the sample was closest to the cluster representing image A, causing the transparency of image A to increase by 5% and the transparency of B to decrease by 5% in the hybrid image seen by the subject. That is, if the proportion of transparency of images A/B was 50%/50% in the previous 100 ms, it would change to 55%/45%. (2) The sample looked more like a sample in the cluster associated with image B, which would lead to a 5% fading in the direction of image B. (3) The outcome was that the sample looked more like images in clusters C or D. This did not result in any change in the hybrid image.

Any one trial could last as little as 1 s (ten consecutive steps from 50%/50% to 100%/0% or 0%/100%). A limit of 10 s was set for each trial, after which the trial was regarded as 'timeout' whatever the transparency of the two images. All the decoding parameters were based on the post-hoc decoding analysis done on a similar MTL population in ref. 14.

Set-up. The experiment was run on a 15-inch laptop computer with images of 160×160 pixels centred on the screen at a distance of about 50 cm from the subject (visual angle of each image of $5.30^\circ \times 5.36^\circ$). Data from the subject's brain was acquired using the Cheetah system (Neuralynx) at 28 kHz, from which it was sent to a server performing spikes detection. Four selected channels were band-pass filtered at 300–3,000 Hz, and a threshold was applied to detect spikes. This threshold was set before the experiment based on a 2-min recording from each channel while the subject was sitting still with eyes opened. Spike counts in the four channels, per 100-ms bin, were transferred via TCP/IP (transmission control protocol/internet protocol) to the experiment laptop computer where the data was used for the online manipulation of the hybrid image. The feedback operation took place in under 100 ms. The experiment was programmed using Matlab (Mathworks) and the Psychophysics toolbox (version 2.54), while the spikes detection proprietary software was written in C++ for efficiency and real-time analysis (code provided on the authors' website at http://www.klab.caltech.edu/~moran/fading).

Response characteristics. We analysed units from the hippocampus, amygdala, entorhinal cortex and parahippocampal cortex. We recorded from 64 microwires in each session. We identified a total of 133 units (68% multi-units and 32% single-units) that were responsive to at least one picture. Out of these responses we selected four in each of 18 sessions. Seven subjects ran one experiment (7×4 units), four subjects ran the experiment twice with two different sets of four units ($4 \times 4 \times 2$ units), and one subject had three sessions, each with a different set of four units ($1 \times 4 \times 3$ units) for a total of 72 units. Out of these responsive units, 58 multi-units and 14 single-units were used in the subsequent fading experiment

(see Supplementary Fig. 2 for a distribution of the units used, and Supplementary Fig. 9 for illustration of the regional competition and performance).

Responses were either positive (exhibiting an increase in the firing rate above baseline, where baseline was determined during the control presentation as described above), or negative (decreasing the firing rate). Excitation was determined using the following techniques developed in previous work[6], by considering the interval after trial onset for all successful trials, divided by the number of spikes. Inhibition was determined using the following four criteria: (1) the median number of spikes in the interval after trial onset for all successful trials, divided by the number of spikes, was at least two standard deviations below the baseline activity, (2) a paired *t*-test using $P = 0.05$ as significance level rejected the null hypothesis of equal means, (3) the median number of spikes during baseline was at least two, (4) the median difference between the number of spikes in the trial and the baseline interval was higher than the background activity of 95 randomly resampled responses (bootstrapping).

Single and multi-units. Spikes used in the analysis were not sorted (that is, clustered) by their shape, but were instead taken as multi-units. This was done to speed up the calculation because template matching of individual spikes on-line had to be sacrificed for the sake of real-time decoding with less than 100 ms delay. Post-hoc analysis of the theoretical performance we could expect had we clustered spikes suggests that it would have increased the performance by 8–10%; however, this is difficult to be sure of because any post-hoc analysis of our data are biased by the fact that we do not have the subjects' feedback to the improved visibility changes on the screen. A further improvement of the set-up would be an additional on-line sorting of spikes, which would lead to a decrease in noise.

Bootstrap testing of statistical significance for task performance. To compare the performance of individual subjects (as in Fig. 3) against chance level we used a bootstrapping technique—generating random trials of activity for each set of four units on the basis of their activity and comparing the mean performance of those to that of the subject. We set individual baselines in the following way: each subjects' sequence of 32 trials (8 trials $\times$ 4 images) was broken into individual 100-ms steps, such that the decoding result for each step was categorized as 'towards target', 'away from target', or 'stay'. For example, in the first trial (coloured red) on the left panel of Fig. 2c (where the target was Marilyn Monroe) the first six 100-ms steps were 'towards target', the seventh 100-ms step was 'towards distractor', the eighth was 'stay', and so on. Thus, each subject ended up having a total number of bins reflecting the proportions of steps he or she used during the course of the entire experiment. This proportion reflected the subject's own baseline chance of going in either direction (the subject in Fig. 2, for instance, had 389 steps where she went towards the target, 49 steps towards the distractor, and 18 'stay' steps altogether). Using these proportions as a priori probabilities, we generated 1,000 new 32-trial blocks. For each 100-ms step, we randomly generated a direction of movement based on the probabilities calculated for each subject, and then generated trials. For each block we calculated the performance and then compared the 1,000 realizations to the one the subject actually performed. If the subject's performance were based only on his/her personal biases (moving in a certain direction because of faster response onset by one unit, paying more attention repeatedly to one of the two competing concepts, and so on) then the random realizations should exhibit a similar performance. The subject's actual performance would be better than the random realizations only if the subject was able to use his or her moves accurately to manoeuvre the fading of the two images towards the target.

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.

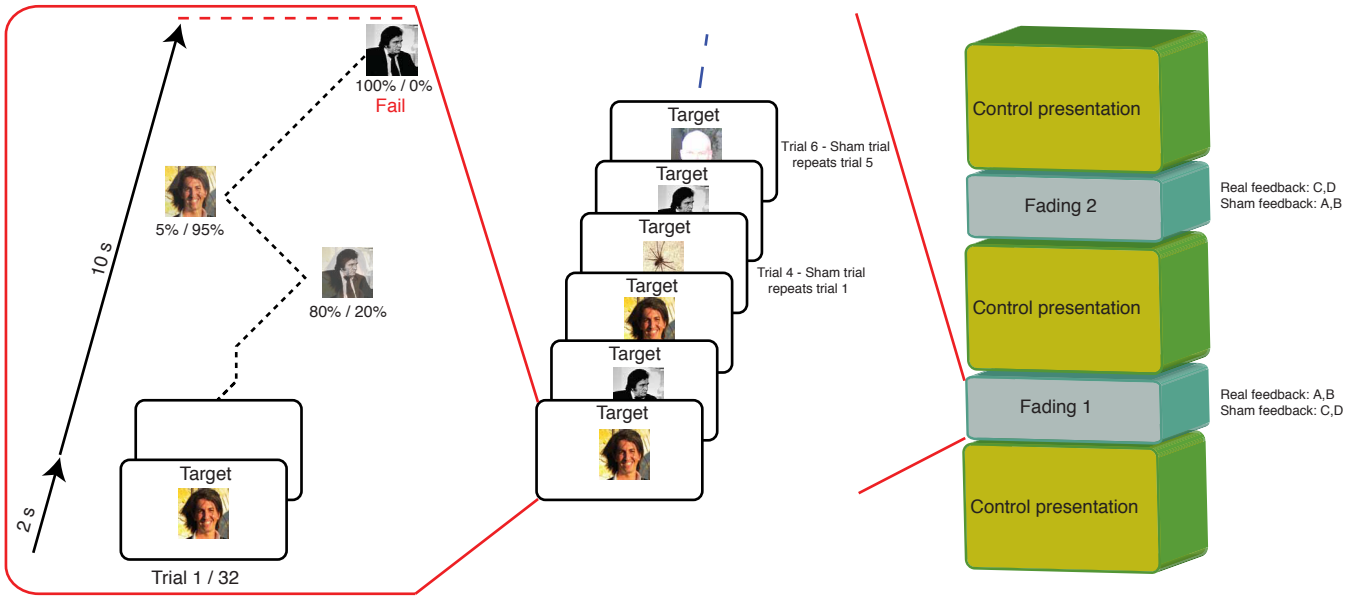

Figure 1. Illustration of the experiment. Right panel shows an illustration of the entire experiment, broken into 5 blocks. The experiment had 3 repetitions of the control presentation (blocks 1, 3, 5), and 2 fading blocks (2 and 4). Real feedback in block 2 was given to two out of the four units, while 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).

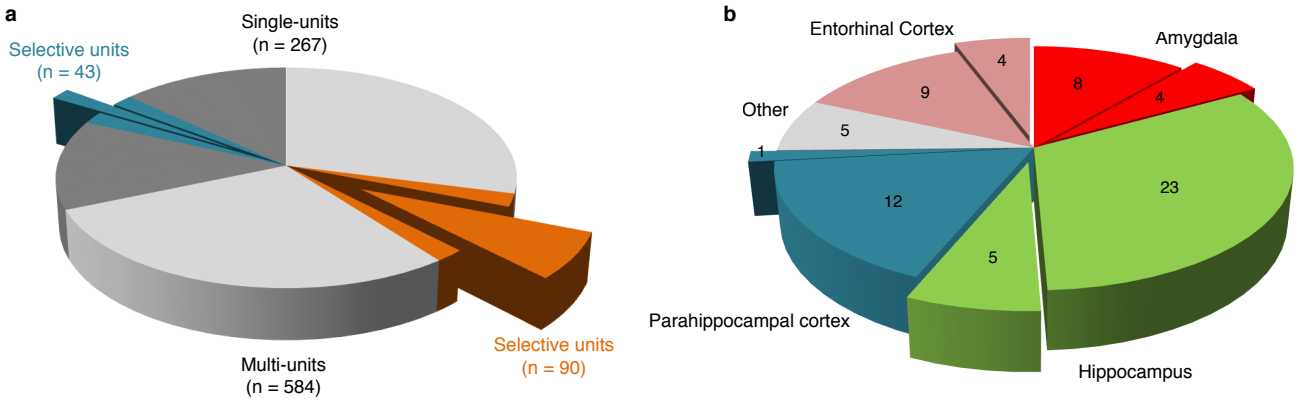

Figure 2. Units distribution. a) Out of 851 units recorded in the course of 18 sessions with 12 subjects, we identified 584 multi-units (69%) and 267 single-units (31%). Out of these, 133 (90 multi-units and 43 single-units) were responsive to one or more images in a prior screening. From these selective 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.

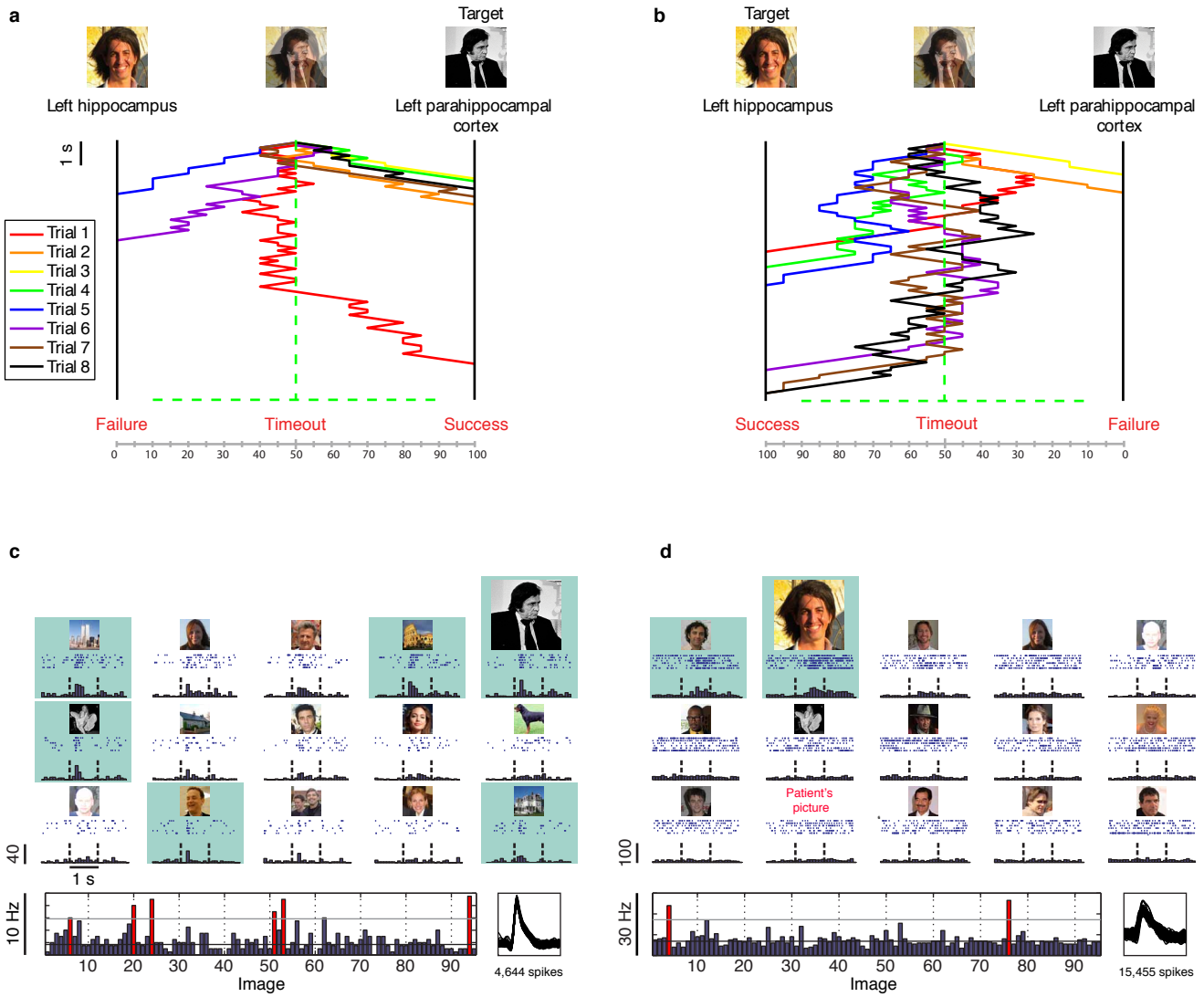

Figure 3. Example of a single block from one subject. Top: Data from 16 trials in which a photo of the first author (with a corresponding neuronal response from a single-unit in the left hippocampus, baseline firing rate: 11.7Hz, firing rate during screening: 25.6Hz, firing rate during fading: 21.0Hz, 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.

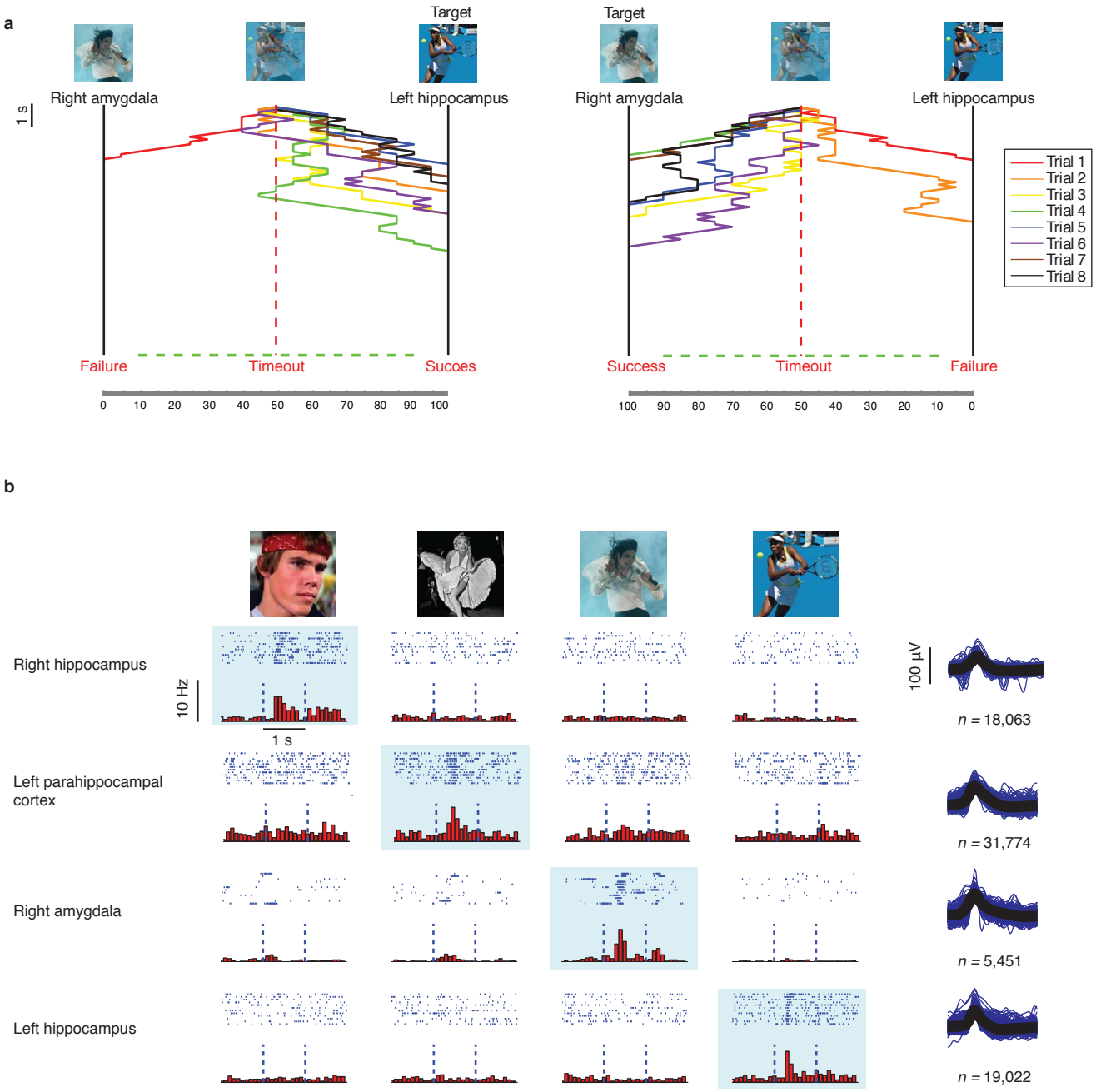

Figure 4. Example of a single session. a) Example of 16 trials for subject 6, where the targets were Michael Jackson (with a responsive single-unit in the subject's right amygdala) and the tennis player Venus Williams (with a responsive multi-unit in the left hippocampus). The subject succeeded in fading 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).

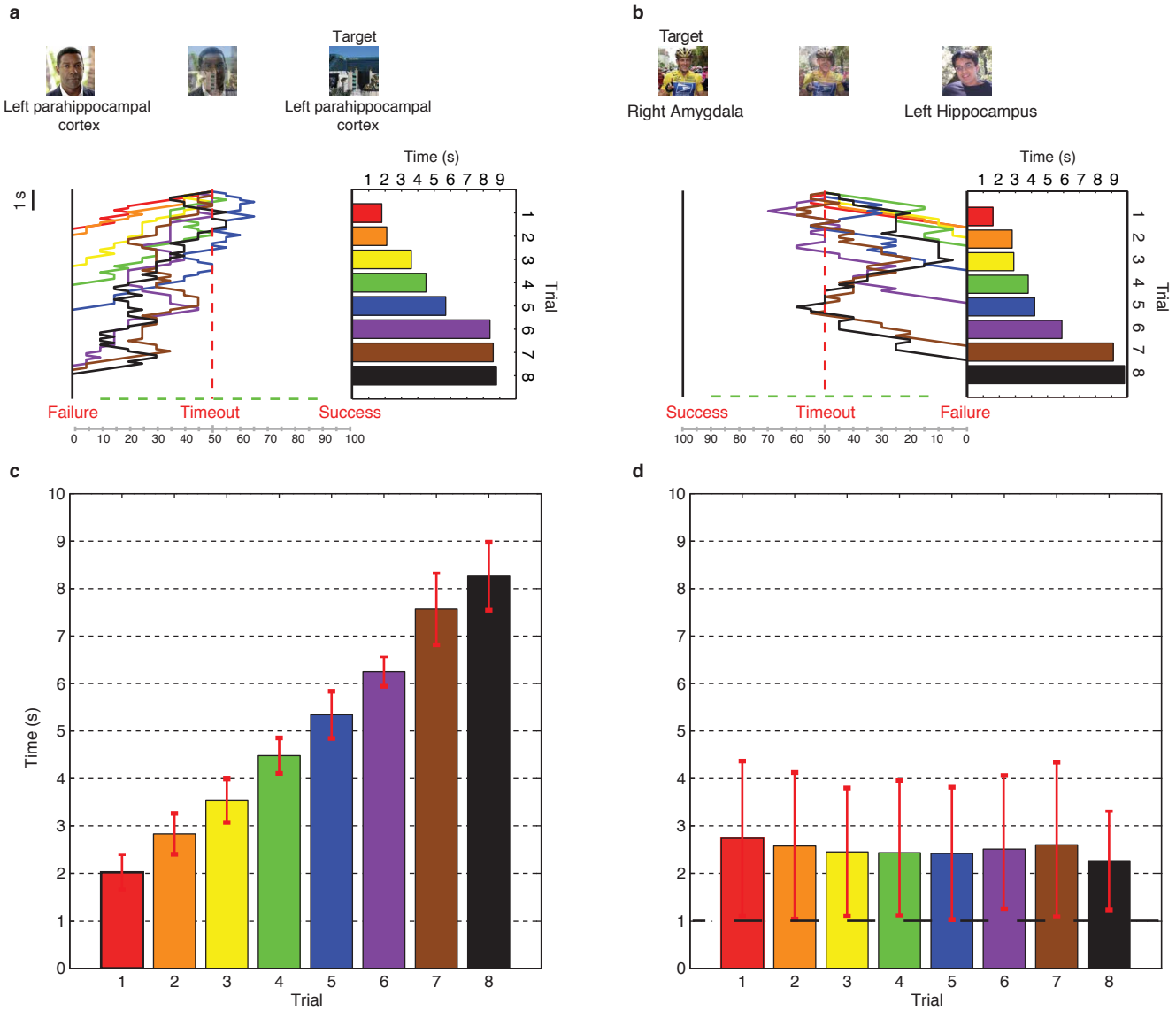

Figure 5. Learning to delay failure. a) Example of learning in a single subject. This subject failed to enhance the image of a building in Las Vegas that the subject was familiar with 8 times. On the right is a corresponding color-coded bar diagram of the time-to-fail - defined as the time to the complete 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.

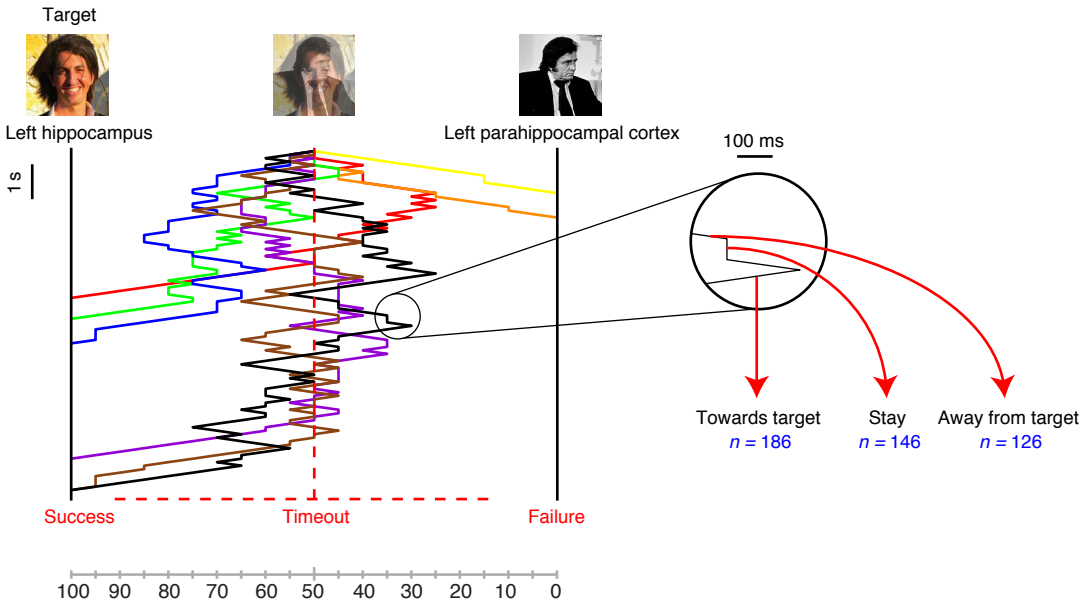

Figure 6. Illustration of the bootstrapping technique. Left panel shows the fading walk for 8 trials in another subject. The firing rate for each unit during fading, divided into 100ms intervals, was classified into one of three categories: 'Towards target' (when the decoding resulted in enhanced visibility 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.

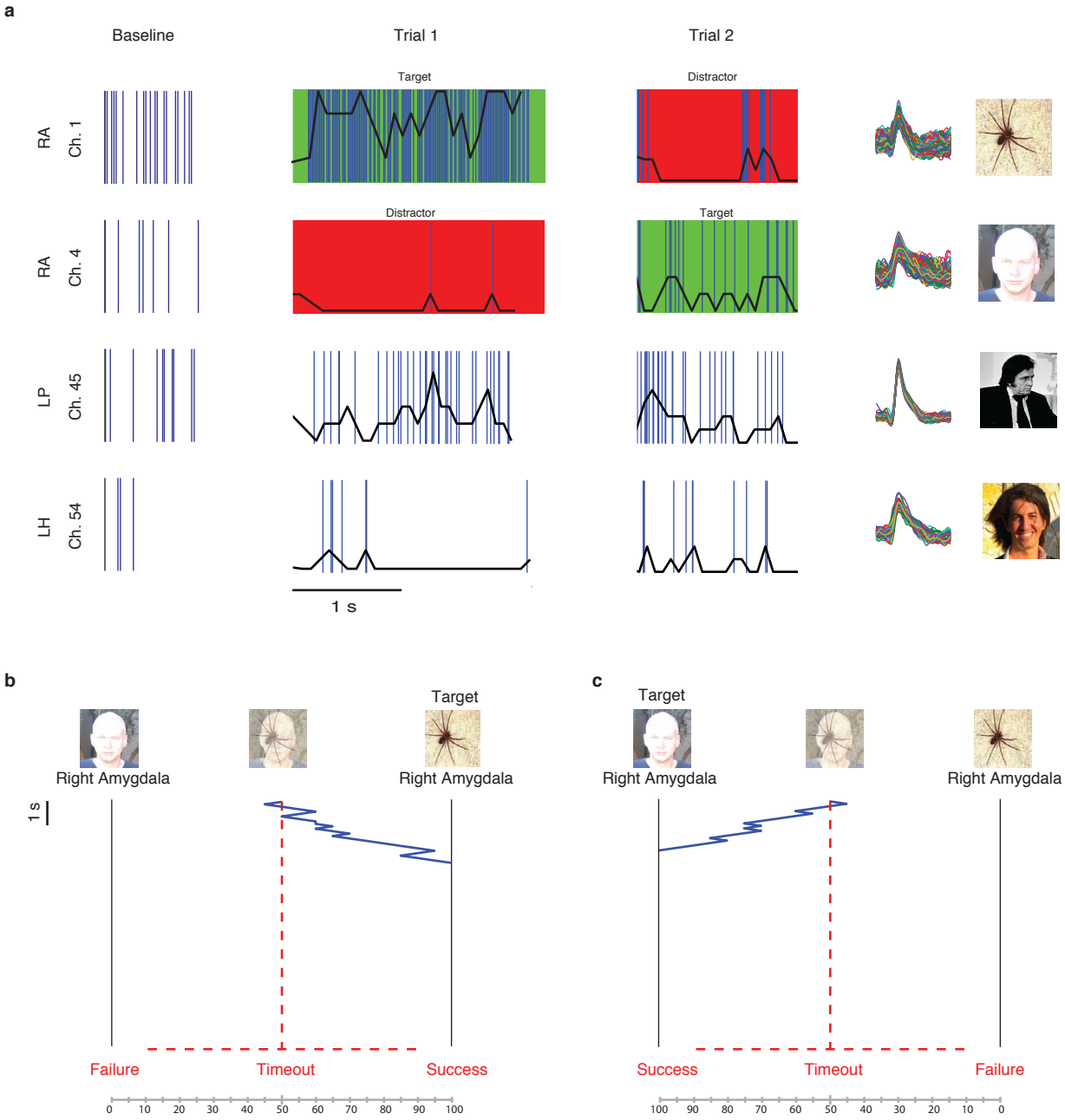

Figure 7. Single trial examples. a) Spiking during the first and the second trials of one subject fading an image of a lab member against an image of a spider. Each row represents an individual unit. Left column illustrates exemplary spikes from a period used for the baseline activity calculation. The 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.

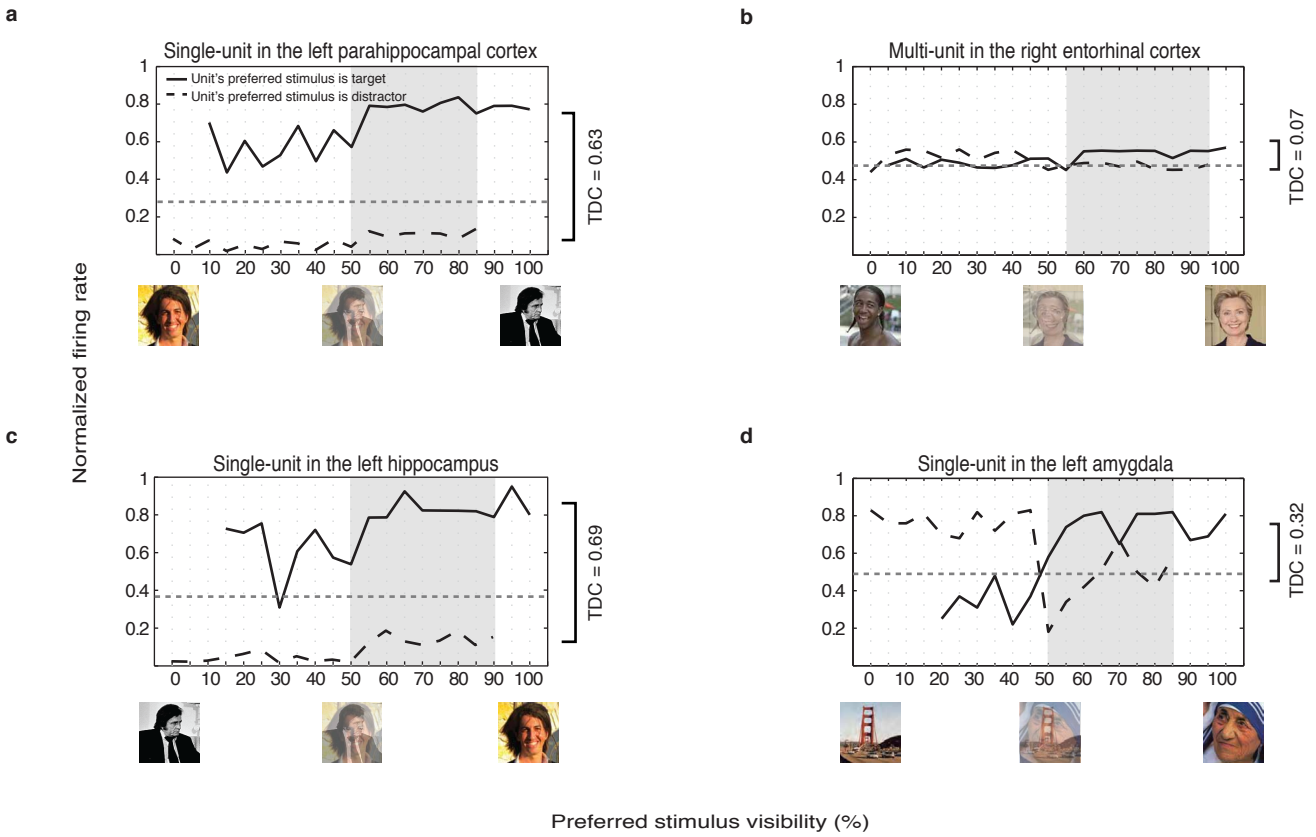

Figure 8. Additional examples of Top-Down-Control. The normalized firing rates of six units in three subjects (panels **a** and **c** are from the subject shown in Supplementary Fig. 3) and sessions as a function of transparency, as in Fig. 4. The top-down-control index for each unit is shown on the 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.

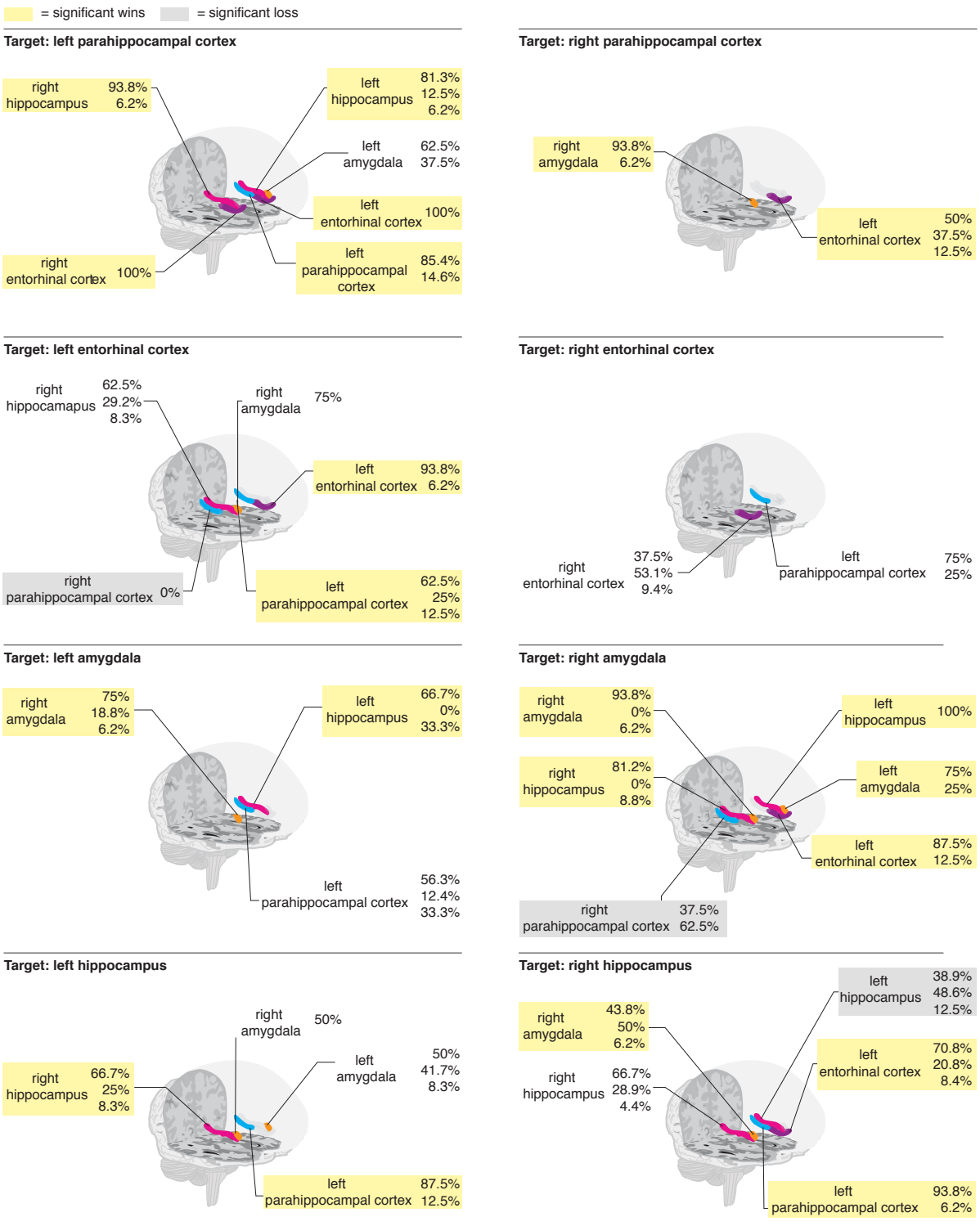

Figure 9. Competition between units across regions. For each of the four regions used in the analysis, we quantified the proportion of trials in which a given unit within a region wins the competition against a unit from a competing region. 'Target' indicates all those units within a specific region 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.*