

On-line, voluntary control of human temporal lobe neurons

Moran Cerf^{1,2,3}, Nikhil Thiruvengadam^{1,4}, Florian Mormann^{1,5}, Alexander Kraskov¹, 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-independent⁵ and invariant manner^{6,7}. These neurons are activated when subjects view⁶, imagine⁸ or recall these concepts or episodes⁹. 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 resection¹³ 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 images⁶. 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 decoder¹⁴ 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

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

*These authors contributed equally to this work.

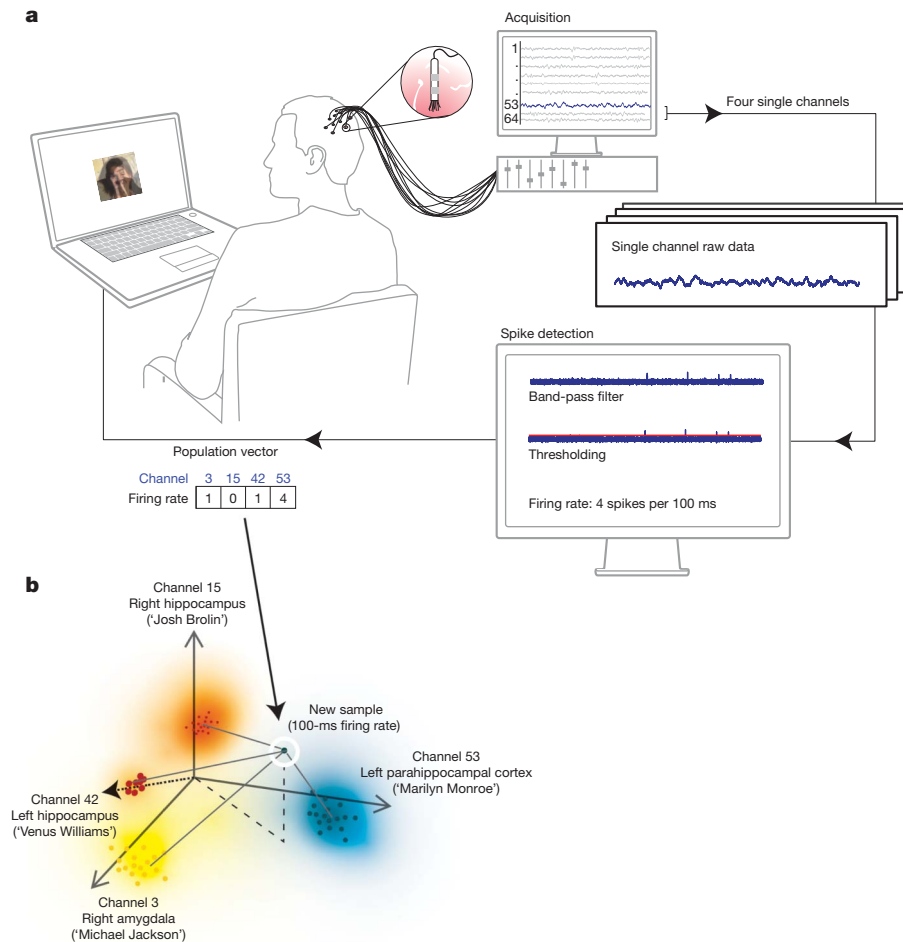


Figure 1 | Experimental set-up. **a**, Continuous voltage traces are recorded by 64 microelectrodes from the subject’s medial temporal lobe. A four-dimensional vector, corresponding to the number of action potentials of four responsive units in the previous 100 ms, is sent to a decoding algorithm determining 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

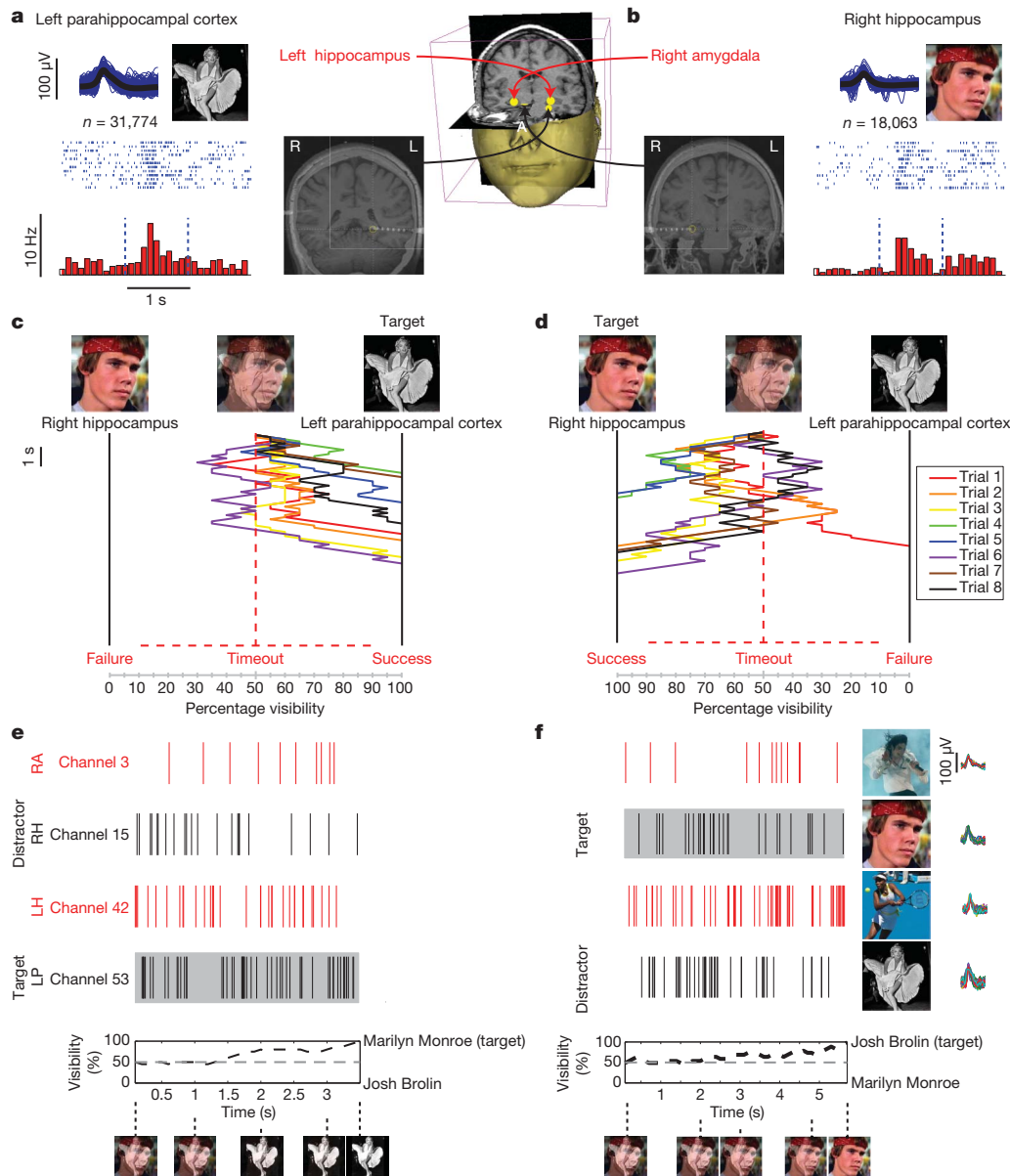


Figure 2 | Task performance and neuronal spiking. Two American actors, ‘Josh Brolin’ and ‘Marilyn Monroe’, constituted the preferred stimulus for two units. **a**, One multi-unit responded selectively to Monroe and was located in the left parahippocampal cortex. Below each illustration are the corresponding 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

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

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.

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

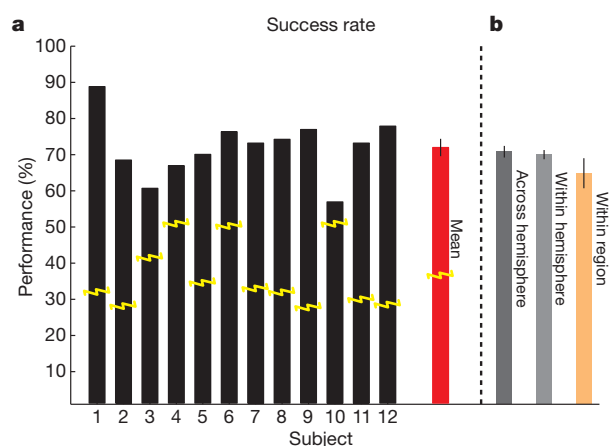


Figure 3 | Successful fading. **a**, Percentage of trials in which subjects successfully controlled the activity of four units and faded to the target image within 10 s. Yellow lines indicate chance performance—determined by bootstrapping 1,000 random trials for each subject ($P < 0.001$; Wilcoxon rank-sum). 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 fields¹³.

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 manner⁵. We previously estimated, using Bayesian reasoning, that any one specific concept is represented by up to one million MTL neurons, but probably by much less²³. As our electrodes are sampling a handful of MTL neurons with predetermined selectivities¹⁴, 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.

- Chalupa, L., Werner, J. & Barnstable, C. *The Visual Neurosciences* (MIT Press, 2004).
- Thorpe, S. Single units and sensation: still just as relevant today. *Perception* **38**, 804–807 (2009).
- Blake, D. A. R. (ed.) *Binocular Rivalry* (MIT Press, 2005).
- Reynolds, J. & Chelazzi, L. Attentional modulation of visual processing. *Annu. Rev. Neurosci.* **27**, 611–648 (2004).
- Quiari Quiroga, R. *et al.* Explicit encoding of multimodal percepts by single neurons in the human brain. *Curr. Biol.* **19**, 1308–1313 (2009).
- Quiari Quiroga, R. *et al.* Invariant visual representation by single neurons in the human brain. *Nature* **435**, 1102–1107 (2005).
- Földiák, P. Neural coding: non-local but explicit and conceptual. *Curr. Biol.* **19**, R904–R906 (2009).
- Kreiman, G., Koch, C. & Fried, I. Imagery neurons in the human brain. *Nature* **408**, 357–361 (2000).
- Gelbard-Sagiv, H. *et al.* Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* **322**, 96–101 (2008).
- Reddy, L., Kanwisher, N. & VanRullen, R. Attention and biased competition in multi-voxel object representations. *Proc. Natl Acad. Sci. USA* **106**, 21447–21452 (2009).
- Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
- Serences, J. *et al.* Control of object-based attention in human cortex. *Cereb. Cortex* **14**, 1346–1357 (2004).
- 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).
- Quiari Quiroga, R. *et al.* Decoding visual inputs from multiple neurons in the human temporal lobe. *J. Neurophysiol.* **98**, 1997–2007 (2007).
- Musallam, S. *et al.* Cognitive control signals for neural prosthetics. *Science* **305**, 258–262 (2004).
- Wessberg, J. *et al.* Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* **408**, 361–365 (2000).
- Velliste, M. *et al.* Cortical control of a prosthetic arm for self-feeding. *Nature* **453**, 1098–1101 (2008).
- Moritz, C., Perlmutter, S. & Fetz, E. Direct control of paralysed muscles by cortical neurons. *Nature* **456**, 639–642 (2008).
- Hochberg, L. *et al.* Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* **442**, 164–171 (2006).
- 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).
- Kennedy, P. *et al.* Direct control of a computer from the human central nervous system. *IEEE Trans. Rehabil. Eng.* **8**, 198–202 (2000).
- Guenther, F. *et al.* A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE* **4**, e8218 (2009).
- 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 criteria¹³. 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 \times 6 repetitions \times (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 \times 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 decoding¹⁴. 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) \times 4 (images) \times 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 data¹⁴, 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:

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

where $\mathbf{x} = (x_1, x_2, x_3, x_4)$ is the new point in the four-dimensional space (corresponding to the firing rate of four units in the previous 100 ms). \mathbf{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 $\bar{\mathbf{S}}$ is the mean of \mathbf{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 work⁶, 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 ($8 \text{ trials} \times 4 \text{ images}$) 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.