Abstract

Social interactions make up to a large extent the prime material of episodic memories. We therefore asked how social signals are coded by neurons in the hippocampus. Human hippocampus is home to neurons representing familiar individuals in an abstract and invariant manner (Quian Quiroga et al. 2009). In contradistinction, activity of rat hippocampal cells is only weakly altered by the presence of other rats (von Heimendahl et al. 2012; Zynyuk et al. 2012). We probed the activity of monkey hippocampal neurons to faces and voices of familiar and unfamiliar individuals (monkeys and humans). Thirty-one percent of neurons recorded without prescreening responded to faces or to voices. Yet responses to faces were more informative about individuals than responses to voices and neuronal responses to facial and vocal identities were not correlated, indicating that in our sample identity information was not conveyed in an invariant manner like in human neurons. Overall, responses displayed by monkey hippocampal neurons were similar to the ones of neurons recorded simultaneously in inferotemporal cortex, whose role in face perception is established. These results demonstrate that the monkey hippocampus participates in the read-out of social information contrary to the rat hippocampus, but possibly lack an explicit conceptual coding of as found in humans.

Key words: area TE, individual recognition, medial temporal lobe (MTL), single-unit, superior temporal sulcus (STS)

Introduction

Individual recognition is achieved by identifying distinct elements such as the face, voice, noun, or silhouette as belonging to one individual. For each well-known individual, a unique combination of these idiosyncratic attributes is represented at the cognitive level. At the neuronal level, this unique representation could arise either from synchronous activation of cell assemblies (Eichenbaum 1993; Hoffman and McNaughton 2002; Canolty et al. 2010) in distant unimodal brain regions (von Kriegstein et al. 2005; von Kriegstein and Giraud 2006) or from the activity of gnostic cells (Konorski 1967) located in a "Person identity node" (Bruce and Young 1986; Burton et al. 1990; Ellis et al. 1997; Campanella and Belin 2007)—although these two hypotheses are not mutually exclusive (Patterson et al. 2007; Meyer and Damasio 2009).

Gnostic cells—cells that responded equally to different pictures of a well-known person’s face, as to its written or pronounced name—have been discovered in human epileptic patients (Quiroga et al. 2005; Waydo et al. 2006; Quian Quiroga et al. 2009). Interestingly, these cells were located in the medial temporal lobe, principally in the hippocampus (but also in the amygdala, entorhinal cortex, and parahippocampal cortex) which is an archaic structure present in all mammals with a highly conserved anatomy, known mainly for its involvement in episodic memory and navigation rather than its role in social information processing (Bird and Burgess 2008; Clayton and Russell 2009; Clark and Squire 2013). It thus raises the questions of (1) whether the hippocampus in different species participates in representing social information (Becker et al. 1999; Ishai et al. 2005; Machado and Bachevalier 2006; Ku et al. 2011; Davidson et al. 2012; Allen...
and Fortin 2013) and (2) whether gnosis neurons representing familiar individuals can be observed in other mammals (Quiroga 2012). Notably, in contradiction with human findings, rats’ hippocampal place cells have been shown to be only weakly sensitive to the presence of a nearby peer (von Heimendahl et al. 2012; Zynuk et al. 2012), but note that in another rodent species hamster’s hippocampal neurons play a role in social odor categorization (Petris et al. 2000, 2005).

In the present study, we aimed at characterizing how hippocampal neurons, recorded without particular prescreening, are activated when monkeys are exposed to different stimuli representing single individuals. To parallel at best the design used in human studies, we presented monkeys with pictures of faces of familiar individuals with different points of view and, in place of nouns, with several audio extracts of voices of these same individuals. We wondered if monkey hippocampal neurons would represent individuals in an invariant and abstract manner similarly to humans’ neurons, if they would on the contrary remain blind to social information similarly to rats’ neurons, or if they would exhibit an intermediate stage of coding. With regard to this issue, rhesus monkeys embody an appealing link between humans and rats. On the one hand, (1) they rely as much on visual information as humans do (Ghaemifar and Santos 2004; Watt and Buchanan-Smith 2006), (2) they individually discriminate (Rendall et al. 1996; Parr et al. 2000; Gathard et al. 2004; Dahl et al. 2007) and recognize (Adachi and Hampton 2011; Sliwa et al. 2011) other monkeys and humans and (3) they possess a rich representation of other individuals encompassing both vocal and facial information (Adachi and Hampton 2011; Sliwa et al. 2011; Habbershon et al. 2013). On the other hand, monkeys lack some semantic and episodic complexity found in humans (Hampton and Schwartz 2004; Bayley et al. 2005; Clark and Squire 2013), leaving the issue of the organization and further of the existence of social concept neurons in their hippocampus open questions.

Finally, we aimed at comparing hippocampal neurons’ activity to the activity of neurons recorded in the inferotemporal cortex (area TE, in the anterior fundus, and anterior lateral part of the lower bank of STS), a region known for its involvement in perceptual processing of faces (Gross et al. 1972; Bruce et al. 1981; Perrett et al. 1982; Rolls 1984; Yamane et al. 1988; Tsao et al. 2006), with the goal of understanding how representation of facial and vocal signals could be transformed along a perception to conceptualization/memory pathway. We therefore additionally recorded neuronal activities from the anterior inferotemporal cortex to the same stimuli and hypothesized that, (1) in both region social stimuli would be categorized separately from nonsocial stimuli at the neuronal and population levels, (2) we would observe a hierarchical processing of information about identity, where the anterior hippocampus would process information about identity independently of the type of modality (similarly to the finding of Quian Quiroga et al. 2009), while inferotemporal neurons would be modality-specific, and (3) seeing or hearing familiar individuals would preferentially trigger hippocampal neurons’ activity as these stimuli might prompt the recall of the whole concept of these individuals or of episodes associated with them (Yanike et al. 2004; Rutishauser et al. 2008; Viskontas et al. 2009).

Materials and Methods

Subjects

All experimental procedures were in accordance with the regulations of local authorities (Direction Départementale des Services Vétéri

Surgical Preparation

Animals were prepared for chronic recording of single-neuron activity in the right hippocampus and right anterior inferotemporal cortex, area TE (Fig. 1A). Anesthesia was induced with Zoletil 20 (15 mg/kg) and maintained under isoflurane (2.5%) during positioning of a clip head-restraint post and recording chamber (Crist Instruments, Damascus, MD). Animals were given atropine (0.25 mg/kg) to prevent excessive salivation. Adequate measures were taken to minimize pain or discomfort. Analgesia was provided by a presurgical buprenorphine injection (0.2 mg/kg). The position of the recording chamber for each animal was calculated using stereotaxic coordinates derived from presurgical anatomical magnetic resonance images (MRI, 0.6 mm isometric), to have access both to the right hippocampus and right TE structure. Post-surgical MR images were used to finely monitor recording locations within the hippocampus and TE during each experiment (Fig. 1A). Structural MRI presurgical and postsurgical scans were performed as animals were anesthetized and placed in an MRI-compatible stereotaxing frame.

Design

To parallel at best the design used in human studies, we presented monkeys with different stimuli representing familiar individuals. Human studies used different pictures of each person as well as their written and pronounced name. Here, we presented monkeys with different pictures of faces of familiar individuals with different points of view and, in place of names, with several audio extracts of voices of these same individuals (Fig. 1B). Because studies carried in humans did not report presenting voices of familiar individuals, it is not known whether the cells responding to faces also respond to the corresponding voices. For example, one can hypothesize that when hearing voices, these cells would have reconfigured to represent meanings of different words rather than identity of the speaker. However, the choice of voices was based on ours and other previous findings that monkeys possess a memory of the association between facial and vocal identities for familiar peers (Adachi and Hampton 2011; Sliwa et al. 2011; Habbershon et al. 2013) and for familiar humans (Sliwa et al. 2011). Importantly, although we used faces and voices, the present study was not aiming at studying multisensory perceptual integration between faces and voices at work, for example, during communication. We rather aimed at testing if a conceptual cross-modal association between face identity and voice identity stored in memory would be represented in an invariant way (F = V) (Stein et al. 2010) at the level of single neurons in the monkey hippocampus. Therefore, we used static pictures of faces and avoided presenting both faces and voices simultaneously because it would have induced the perception of congruency in stimuli dynamics and not only congruency in identity. As control stimuli, we presented to the monkeys pictures and acoustic extracts of unknown individuals. We also presented visual and acoustic objects and synthetic patterns, which have been shown to elicit response activity in hippocampal
neurons (Tamura, Ono, Fukuda and Nakamura 1992; Tamura, Ono, Fukuda and Nishijo 1992; Rolls et al. 1993; Hampson et al. 2004; Rolls et al. 2005). Finally, contrary to the studies conducted by Quiroga et al., we did not attempt to prescreen neurons, giving us the chance to document responses to facial and vocal stimuli along a large spectrum of selective activities, and not only in ultra-selective cells.

Behavioral Procedure

The animal was head restrained and placed in front of, and at eye level to, an LCD screen situated at 56 cm, in a quiet room. Adequate measures were taken to minimize pain or discomfort. The subject’s gaze position was monitored with an infrared eye-tracker (ISCAN) with a 250-Hz sampling rate. Behavioral paradigms, visual displays, eye position monitoring, and reward delivery were under the control of a computer running a real-time data acquisition system (REX) (Hays et al. 1982). Visual stimuli subtended a visual angle of 10° × 13° on the center of the screen (Fig. 1C). The virtual exploration window consisted of the picture surrounded by ~2° black surround visual angle. Auditory stimuli were presented in a quiet room at the intensity of a regular conversation, that is, 50–65 dB (A-weighted) sound pressure level (SPL) at the subject’s ear as measured with a Brüel and Kjær 2239A Integrating Sound Level Meter (http://www.bksv.com) from 2 speakers located at 56 cm in front of the subject and symmetrically 45 cm apart. To start the trial, the subject directed its gaze to a fixation point at the center of the screen. Then it was required to fixate (within ±4°) during 0.5 s. Two types of test trials were randomly interleaved: trials with visual stimuli and trials with auditory stimuli (Fig. 1C). In visual trials, after the 0.5 s fixation, an image was presented at the center of the screen. The subject could freely explore the picture during 2 s, so long as its gaze was maintained within the boundaries of a virtual window around the picture corresponding to the black area in Figure 1B. In auditory trials, after the 0.5 s fixation, the subject was required to continue fixating (+4°) during 2 s while an audio sample was played. Juice reward was given after each trial to ensure monkeys’ motivation to complete trials and their attention to the computer monitor where visual stimuli were displayed. Thus reinforcement did not depend on any particular exploration pattern.

Stimuli

Visual stimuli were color images of individuals (humans and rhesus monkeys), objects and synthetic pictures (Fractal Explorer, fractals.da.ru). Because our focus was to investigate a code for identity, we presented more stimuli of individuals than of objects, allowing testing for neuronal activity in response to social and perceptual dimensions of face stimuli (gender, view, and species). Thus, there were three pictures (~30°/30°/front view) for each of the 6 human individuals and 6 or 5 monkey individuals shown and one picture per object and fractal; leading to 18 human, 18 or 21 monkey, 6 object and 3 fractal pictures presented to each tested animal (Fig. 1B). Visual stimuli were presented on a black background. Face photographs were cropped to only include the face, head, and neck, and then adjusted to a common size (GNU Image Manipulation Program, http://www.gimp.org). Human faces were presented without masks and
stimulus set was specified, in order to avoid a novelty confound (Xiang and Brown samples. These stimuli have been presented to the monkeys on duals and objects, as well as fractal images and synthetic audio iarity with individuals presented in stimuli. Unfamiliar stimuli present in the colony rooms (hose, primate chair, and pole). In during the same 2 to 3-year period. Familiar objects were objects prior to testing. Familiar human individuals were a care-giver c a q u e sh o u d be int h es a m er o o ma sth es u b j e c ts f o r2 iar. Familiar simian individuals were the 2 or 3 adult rhesus ma- familiar to the subjects (including the subject itself), or unfamil- iar. Familiar simian individuals were the 2 or 3 adult rhesus ma-

goggles (that they wore in the presence of the animals) since the colony rooms were equipped with windows through which animals could see the experimenters and staff without this protection. Pictures shown in the document are provided for illustrative pur-

This set was mirrored by an acoustic set. Three audio samples from each of the 6 human individuals and 6 or 5 monkey individ-

Individuals and objects represented on the stimuli were either familiar to the subjects (including the subject itself), or unfamiliar. Familiar simian individuals were the 2 or 3 adult rhesus macaques housed in the same room as the subjects for 2–3 years prior to testing. Familiar human individuals were a care-giver and 2 experimenters working with the animals on a daily basis during the same 2 to 3-year period. Familiar objects were objects present in the colony rooms (hose, primate chair, and pole). In the manuscript, these stimuli are referred to as “known.” Each stimulus set was specific to each animal to ensure a high famil-

Individuals and objects represented on the stimuli were either familiar to the subjects (including the subject itself), or unfamiliar. Familiar simian individuals were the 2 or 3 adult rhesus macaques housed in the same room as the subjects for 2–3 years prior to testing. Familiar human individuals were a care-giver and 2 experimenters working with the animals on a daily basis during the same 2 to 3-year period. Familiar objects were objects present in the colony rooms (hose, primate chair, and pole). In the manuscript, these stimuli are referred to as “known.” Each stimulus set was specific to each animal to ensure a high famil-

Electrophysiological Recordings

Single-neuron activity was recorded extracellularly with tungsten quartz insulated tetrodes (1.5–2 MΩ; Thomas Recording, Giessen, Germany) or tungsten single microelectrodes (1–2 MΩ; Frederick Haer Company, Bowdoinham, ME) which were lowered into the hippocampus and area TE. The depth of recording was calculated based on the distance between the target area and the tip of a reference electrode that was viewed on the MRI slice and served as a reference (Fig. 1A). The microelectrodes were inserted through 24-gauge stainless steel guide tubes set inside a delrin grid (Crist Instruments) adapted to the recording chamber. The electrodes were lowered individually with a NAN microdrive (Plexon Inc., Dallas, TX) (Fig. 1A) at a speed of 20–30 μm/s to the required depth under electrophysiological monitoring; background activity and neural signals changed at the crossing of boundaries between cortex, white matter, and gray matter at the expected depths. For the hippocampus we also used the signal change specific to the ventricles as a land-

Cells Location

Recordings were carried in the anterior part of the right hippo-

campus, mainly between 15 and 18 mm on the anteroposterior axis (AP, relative to the interaural line) in one monkey and 8–16 mm in the other one (Fig. 1A and Supplementary Fig. S1). Recordings were carried out in the superior part of right TE, that is, in the interior inferior bank of STS, mainly in its anterior fundus and anterior lateral part between 10–12 mm in one mon-

Cells Selectivity Analysis

Data were analyzed with custom written scripts and the Statistics Toolbox in Matlab (MathWorks). Neurons with very low firing rates (FRs) (<0.4 Hz) were not included in the following analysis due to their lack of reliability when tested with the following methods. Raster plots and peristimulus histograms in response to the stimuli set were plotted for each neuron for inspection. The baseline activity was defined as the average activity during 500 ms preceding the stimulus onset during fixation. The peak of each neuron’s activity was defined separately for auditory and visual modalities as the extremum of their trial-averaged
firing rate calculated on 50 ms consecutive bins. Then neuron’s response to a given stimulus was defined as the trial-averaged FR in a 500-ms window centered on the time of the peak stimulus activity. This method was used to take into account the use of dynamic acoustic stimuli, which can lead to different response time courses among different neurons. In parallel, spike trains were also smoothed by convolution with a Gaussian kernel (σ = 10 ms) to obtain the spike density function (SDF). A neuron’s response was considered excitatory if it was significantly different from baseline activity (P < 0.05, t-test) and if its SDF was greater than the mean plus 4 standard deviations (SDs) of the baseline activity. A neuron’s response was considered inhibitory if it was significantly suppressed for at least 100 consecutive milliseconds during a window of 500 ms after the stimulus onset. Neurons were classified according to their responses to either or both of the 2 modalities.

We then wondered if, and how many, neurons would respond to social stimuli compared with nonsocial stimuli in either one of both modalities. Neurons were defined as face-, object-, voice-, or sound-selective if their activity to at least one stimulus of these categories was significant and if additionally their activity was not significant for stimuli of the other category from the same modality. We used a stringent criterion (4 STD above baseline) because we did not have an equal number of faces and objects in our stimulus set. This criterion differs from previous reports describing activity to faces in which responses to face stimuli were 2–10 times larger than responses to objects (Rolls 1984).

Second, stimuli represented either known or unknown individuals and also varied along several other categories: viewpoint (–30°/50°/frontal), species (human/monkey), gender (female/male) and identity. Therefore, to determine whether these categories are preferentially driving face-selective cells’ activity, a generalized linear model (GLM) with 4 factors (familiarity, species, gender, and viewpoint) was fitted on each face-selective neuron response to the visual subset of stimuli. Similarly, a GLM with three factors (familiarity, species, and gender) was fitted on each voice-selective cell’s response to the acoustic subset of stimuli. Only selective neurons presenting FR increases (but not decreases) for their preferred stimuli were included in this and following analyses. To assess if stimuli of familiar (rather than of unfamiliar) individuals are driving neurons’ activity, we calculated the number of familiar images/sounds generating a response in each face/voice-selective cells (Viskontas et al. 2009) and compared it with a paired Student test to the number of unfamiliar images/sounds in the same cells.

Third, we quantified with a receiver operating characteristics (ROC) analysis, if neuronal activity related to facial (or vocal) identity was invariant, that is, if the average FRs to the three viewpoints (or vocal extracts) of a given person were similar. The hit rate corresponded to the median number of spikes pulled across all trials for each of the three pictures (or audio extracts) of an individual, and the false-positive rates to the median number of spikes pulled across all trials elicited by each of the other visual (or acoustic) stimuli of the set. Ninety-five percent confidence intervals were estimated using a bootstrap of the stimuli responses. The ROC analysis was performed for each facial (or vocal) identity and the maximum area under the curve (AUC) from all of them was designated as the best ROC area (Eifuku et al. 2011). In parallel, we also asked how selective each neuron’s responses were, by quantifying the number of stimuli eliciting activity greater than the half-maximum activity of this neuron (Perrodin et al. 2011) and also by computing a sparseness index $S$ (Rolls and Tovee 1995), with $r_i$ being the response of the neuron to the $i$th stimulus, and $N$ being the total number of stimuli, as follows:

$$S = \frac{\sum_{i=1}^{N} r_i/N}{\sum_{i=1}^{N} r_i^2/N}$$

The sparseness index ranges from 0 (highly-selective or sparse coding) to 1 (nonselective or dense coding).

Fourth, face–voice association at the neuronal level was investigated. For each cell the correlation coefficient between the response vector to the different facial identities and the response vector to corresponding vocal identities was calculated. To determine if the calculated correlation coefficient could have been obtained by chance, we compared them with correlation coefficients calculated by permuting identities 719 times.

Finally, unsupervised methods were also used to assess finer selectivity. For each neuron a principal component analysis was performed, with observations being each presented stimulus from the set and variables being the neuron’s response at each time bin for the respective stimuli. Stimuli responses were projected on the first 2 principal components and then clustered either using a Gaussian mixture model or by calculating the Euclidian distance between the representation of the neuronal responses to every pair of stimuli in the principal component space and further linking them hierarchically into a cluster tree.

**Population Analysis and Comparison Across Brain Areas**

$\chi^2$ tests were used to compare across brain regions the number of cells responding to the different modalities and categories. We also compared the indexes of selectivity across modalities and across regions with Student and Wilcoxon tests. Finally unsupervised methods were used to determine which parameters are driving neuronal activity (of all the recorded neurons and not only of the selective ones) at the population level in both brain areas. For each given stimulus and each given neuron, a z-score was calculated as the neuron’s mean FR for the stimulus in a 500 ms window centered at the peak of activity for this stimulus divided by the standard deviation of the FR in the same 500 ms window. The peak of activity was defined as the extremum of the trial-averaged FR on 50 ms consecutive bins for a given stimulus (if the activity was similar to that of the baseline, the z-score was close to 1/SD of the FR in the 500-ms window). The z-scores weighted the contribution of change in FR in response to a stimulus presentation by its reliability across trials, whereby less reliable responses (i.e., with a high standard deviation) were weighted down compared with more reliable responses (i.e., with a low standard deviation). Principal component analyses on the z-scores were performed in both the hippocampus and TE, with observations being each one of the presented stimuli and variables being each neuron. Z-scores between the different neurons were also standardized in order for the PCA not to be dominated by neurons with the highest modulation depths (Cunningham and Yu 2014). The standardized z-scores were projected on the first 2 principal components and then clustered using a Gaussian mixture model, which is based on an expectation–maximization (EM) algorithm. A hierarchical analysis was also performed, by calculating the Euclidian distances in the principal component space between every pair of stimuli and further linking the stimuli hierarchically into a cluster tree. This analysis assumes the existence of some categorical structure, contrary to the unsupervised stimulus arrangements done with
the PCA, but it does not assume any particular grouping into categories (Kriegeskorte et al. 2008).

Results

The activity of 343 neurons was recorded in 2 monkeys in the hippocampus (188 cells: 99 in monkey O and 89 in monkey Y; Fig. 1A; 58 in CA1, 110 in CA3, 18 in the dentate gyrus and 2 in the subiculum) and in area TE (155 cells: 82 in monkey O and 73 in monkey Y; Fig. 1A), while monkeys were presented color images or audio-recordings of individuals and objects familiar to them (including themselves) or not personally familiar (Fig. 1B). Spontaneous FR of cells ranged from 0.1 to 29 Hz in the hippocampus (median 1.3 Hz, mean 3.2 Hz) and from 0.1 to 127 Hz in area TE (median: 1.4 Hz, mean: 3.8 Hz). Thus, the population of cells recorded is likely to encompass pyramidal cells and interneurons in the hippocampus (Matsumura et al. 1999; Wirth et al. 2003; Viskontas et al. 2007; Ison et al. 2011) and area TE (Mruzek and Sheinberg 2012). Naturalistic stimuli elicited robust responses throughout the sampled regions, with 83 (44%) hippocampal neurons (26 in CA1, 50 in CA3, Fig. 2A) selective to at least one stimulus in either modality (with criteria: excitatory response, $P < 0.05$, median FR superior to 4 SD of the baseline and superior to 0.4 Hz) and 79 (51%) neurons in TE (Fig. 2A). In the hippocampus, 11.7% of cells showed activity that was significantly suppressed by the visual stimulus onset relative to baseline. In contrast, 25.5% of cells showed excitatory responses to the stimulus onset and 7% showed mixed responses (excitatory to some stimuli and inhibitory to others). Inhibitory responses to acoustic cues were less frequent: only 5% were inhibitions, 2% showed mixed inhibitory and excitatory responses, and 23% were excitatory responses. The same pattern was found in TE for visual cues. Thirty-seven (24%) inferotemporal neurons were found voice-selective (Fig. 2B and Supplementary Figs S10, S11). Conversely, 5 (3%) were sound-selective. These proportions also did not differ between TE and the hippocampus and did not differ compared with the proportion of face-selective cells in TE ($P > 0.05$, $\chi^2$-test). Thirty-seven (24%) inferotemporal neurons were found voice-selective (Fig. 2B and Supplementary Figs S10, S11). Conversely, 5 (3%) were sound-selective. These proportions also did not differ between TE and the hippocampus and did not differ compared with the proportion of face-selective cells in TE ($P > 0.05$, $\chi^2$-test). Compared with face-selective cells, voice-selective cells displayed less pronounced increase in FRs compared with their baseline activity (2-sided t-test, $P = 0.028$ Fig. 2C). Of the 155 cells recorded in TE, 32 (21%) were found to be face-selective (Fig. 2B and Supplementary Fig. S11A–C). Example face-selective cells from TE are presented in Figure 2D and Supplementary Figures S8, S9. Conversely, 10 (6%) were object-selective. These proportions did not differ from those found in the hippocampus ($P > 0.05$, $\chi^2$-test). Thirty-seven (24%) inferotemporal neurons were found voice-selective (Fig. 2B and Supplementary Figs S10, S11). Conversely, 5 (3%) were sound-selective. These proportions also did not differ between TE and the hippocampus and did not differ compared with the proportion of face-selective cells in TE ($P > 0.05$, $\chi^2$-test). Compared with face-selective cells, voice-selective cells displayed less pronounced increase in FRs compared with their baseline activity (2-sided t-test, $P = 0.02$ in TE, as was observed in the hippocampus. These first results show the existence of cells in the monkey hippocampus, similar to cells in TE, which respond differentially to social stimuli compared with nonsocial stimuli in the absence of any conditional training of the animals. These results suggest that acoustic signal might be less well categorized into social and nonsocial than visual signals in the monkey hippocampus and in TE.

Representation of Social Categories

To determine which categories preferentially drive the activity of face-selective cells, a generalized linear model GLM analysis investigated the selectivity of these cells with 4 factors characterizing stimuli in the experimental dataset: (1) familiarity (known/unknown), (2) species (monkey/human), (3) gender (female/male), (4) viewpoint (frontal/30°/–30°). We found that, 11 (29%) of the 38 hippocampal neurons responded in a differential fashion to at least one of these 4 factors (Table 1). The species factor characterized the largest proportion of cells for which activity was significantly explained by the GLM (Table 1, Fig. 2C,D, and Supplementary Figs S2, S3 for single examples of cells responding more to monkey faces).

In TE, 14 (44%) of the 32 TE neurons were found to respond in a differential fashion to at least one of the 4 factors (familiarity, species, viewpoint, and gender, Table 1). As in the hippocampus, the species factor characterized the largest proportion of cells for which activity was significantly explained by the GLM (Table 1, Fig. 2D, and Supplementary Figs S8, S9 for single examples of cells responding more to human or monkey faces, respectively). In this case the proportion of responses to the 4 factors as a whole differed significantly between the hippocampus and TE ($P = 0.0022$, $\chi^2$-test); inferotemporal cells being overall more sensitive to coding facial category that hippocampal cells.

A GLM analysis was also used to investigate the selectivity of voice-selective cells with three factors characterizing stimuli in the experimental dataset: (1) familiarity (known/unknown), (2) species (monkey/human), (3) gender (female/male). We found that only 4 (14%) of the 28 hippocampal cells responded in a differential fashion to at least one of these three factors (Table 1). These proportions were smaller than the proportions observed for face-selective cells, though not significantly ($P = 0.29$, $\chi^2$-test). In TE, 4 (11%) of the 37 neurons responded in a differential fashion to at least one of the three factors (familiarity, species, gender, Table 1). These proportions were significantly different from those in the hippocampus.
smaller than the proportions observed for face-selective cells in the same structure ($P = 2.10^{-9}$, $\chi^2$-test), showing that in the inferotemporal cortex, responses to voices are less informative than responses to faces as was also the case in the hippocampus.

**Representation of Familiarity**

To assess if stimuli of known (rather than of unknown/not personally familiar) individuals are driving neurons’ activity, we calculated the number of known images/sounds generating a response in face/voice-selective cells (Viskontas et al. 2009). In the hippocampus, there were $2.13 \pm 0.37$ pictures of known individuals eliciting a selective response in face-selective neurons, compared with $2 \pm 0.32$ pictures of unknown individuals. There were $0.64 \pm 0.16$ sounds of known individuals eliciting a selective response in voice-selective neurons, compared with $0.71 \pm 0.21$ sounds of unknown individuals. These numbers did not differ significantly ($P = 0.68$, $P = 0.77$, paired t-tests) showing that known individuals were not preferentially coded compared with unknown individuals by hippocampal neurons in the...
monkey. In TE, 2.03 ± 0.4 pictures of familiar individuals elicited a selective response in face-selective neurons, compared with 2.31 ± 0.37 pictures of unfamiliar individuals. There were 0.78 ± 0.12 sounds of familiar individuals that elicited a selective response in voice-selective neurons, compared with 0.84 ± 0.17 sounds of unfamiliar individuals. These amounts did not differ significantly (P = 0.29, P = 0.82, paired t-tests) showing that familiar individuals were not preferentially coded compared with unfamiliar individuals by inferotemporal neurons, as was observed in the hippocampus (P = 1, χ²-test). Note that both stimuli of known and unknown individuals and objects have been seen/heard by the monkeys hundreds of times prior to the recording sessions (both stimuli being thus visually or acoustically familiar). Therefore, the results shown above do not generalize to novel images, for example, shown maximally a few times in a lifetime, which we have not presented in this study, and only concerns stimuli of unfamiliar (in the sense not personally familiar) individuals presented several times prior to the experiment.

Response to Identities

Binary classifiers tested for face-selective cells’ invariant coding of facial identity throughout viewpoint, compared with the rest of the visual stimuli. For each face-selective cell, the best AUC of the 12 ROC curves was calculated (see Supplementary Figs S2C and S8C for examples). The greater the AUC, the higher a cell discriminates between the three viewpoints of one individual against other stimuli, and thus the higher is cell invariance for facial identity. In the hippocampus, the mean invariance of face-selective cells activity to facial identities was AUCbest,HPC = 0.83, (Fig. 3A top) which did not differ significantly from the mean invariance of face-selective cells in TE (AUCbest,TE = 0.86, P > 0.05, 2-sided Wilcoxon rank sum test, Fig. 3A). The mean invariance of voice-selective cells activity to vocal identities was calculated (see Supplementary Figs S6C and S10C,D for examples) and was found to be lower compared with face-selective cells identity invariance though not significantly in hippocampus (AUCbest,HPC = 0.80, Fig. 3A top). In TE, the mean invariance of voice-selective cells activity to vocal identities was found to be significantly lower compared with face-selective cells invariance to facial identities (AUCbest,TE = 0.77, P = 0.006, 2-sided Wilcoxon test, Fig. 3A bottom). Even though many cells coded for the different facial views of at least one individual in a similar fashion, these cells did not necessarily respond to a unique individual. On the contrary, an analysis of the sparseness of the cells (i.e., whether cells were active for most stimuli or for few stimuli) showed that only very few cells (e.g., Supplementary Fig. S5) displayed high selectivity, in both areas (see below).

Sparerness of Face-selective and Voice-selective Cells Activity

The selectivity of face and voice-selective cells was first compared by quantifying the number of stimuli eliciting an activity greater than the half-maximum activity of that cell (Perrodin et al. 2011). In the hippocampus, face-selective cells responded to an average of 51% of the visual stimuli, that is, 23/45 of the stimuli elicited responses greater than the half-maximum response (Fig. 4A,D). Similarly to the hippocampus, inferotemporal face-selective cells responded to an average of 42% of the visual stimuli, that is, 19/45 of the visual stimuli elicited responses greater than the half-maximum response (P = 0.13, Wilcoxon test, Fig. 4A,D).

In the hippocampus voice-selective cells presented a higher selectivity than face-selective cells (average percentage of acoustic stimuli eliciting a response: 34%, that is, 15/45 stimuli, P = 0.0065, Wilcoxon test, Fig. 4E,H). In contrast in TE, the average number of acoustic stimuli eliciting a response was similar to the average number of visual stimuli eliciting a response (45%, i.e., 20/45 stimuli, P = 0.054, Wilcoxon test, Fig. 4E,H) and thus significantly higher than in the hippocampus (P = 0.049, Wilcoxon test, Fig. 4E).

We further analyzed whether face- and voice-selective cells were active for most stimuli or for few stimuli by computing a sparseness index, which ranges from 0 (highly-selective or sparse coding) to 1 (non-selective or dense coding) and obtained the same results. In the hippocampus, the mean sparseness index was high for both types of selective cells (0.83 for face-selective and 0.77 for voice-selective; Fig. 4B,F) indicating dense coding yet sparser for vocal stimuli than for face stimuli (P = 0.030, Student test). Face selectivity did not differ between CA1 and CA3 (0.82 in CA1, 0.84 in CA3, P > 0.05, Student test; Fig. 4C); nor did selectivity for voices (0.81 in CA1, 0.74 in CA3, P > 0.05 Student test; Fig. 4G). In TE, the mean sparseness indexes were high for both face- and voice-selective cells (0.80, 0.81 respectively, P > 0.05, Fig. 4B,F).

Both indexes were similar across modalities and across regions; indicating a dense coding in the hippocampus as was observed in TE—with only a sparser coding for vocal stimuli compared with facial ones in the hippocampus. On the contrary these indexes were higher than those described in some human studies in the hippocampus (Quiroga et al. 2005) and ultra-sparse hippocampal cells were only anecdotally (<1%) observed in our study (e.g., Supplementary Fig. S5).

Face-voice Association

We investigated whether cells encoded identity throughout modalities presentation. Some hippocampal neurons were activated by stimuli from both modalities (“bimodal cells”) (Fig. 2A). 23 (12%) cells were bimodal, compared with 38 (20%) visual and 22 (12%) auditory cells. From these cells, 7 responded to both faces and voices and not to sounds and objects stimuli (Fig. 2B). These latter cells were analyzed in search of activity invariant to modality. We found that these cells’ responses to facial identities were poorly correlated to their corresponding vocal identities (r = 0.22, P = 0.22, Student test as compared with a distribution of correlations obtained by chance through permutations, Fig. 3B top). Figure 3C also shows that the distribution of

Table 1 Face-selective and voice-selective cells properties in the hippocampus and TE

<table>
<thead>
<tr>
<th></th>
<th>Percentage of face-selective cells</th>
<th>Percentage of voice-selective cells</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HPC</td>
<td>TE</td>
</tr>
<tr>
<td>Known–unknown</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Monkey-human</td>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>Female–male</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>View (0°, 30°, and –30°)</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Not modulated</td>
<td>71</td>
<td>56</td>
</tr>
</tbody>
</table>

Note: Percentage of cells modulated by one of the 4/3 categories as assessed with generalized linear model analyses carried separately on each population.
the responses to the best familiar face compared with the corresponding voice were biased towards the face response for face-selective neurons while the opposite pattern was found for voice-selective neurons (see Supplementary Fig. S11D for responses to unfamiliar individuals). Thus it appears that although some cells respond to multiple views of the same individual (e.g., Supplementary Figs S2 and S3), this invariance does not generalize to voice stimuli of the same individual.

In TE, 32 (21%) cells were bimodal, compared with 34 (22%) visual and 13 (8%) auditory cells (Fig. 2A). From these bimodal cells, 14 responded to both faces and voices and not to sounds and objects stimuli (Fig. 2B). These latter cells were analyzed in search of activity invariant to modality. We found that the cells’ responses to facial identities were poorly correlated to their corresponding vocal identities ($\rho = -0.07$, $P = 0.61$, Student test as compared with a distribution of correlations obtained by chance through permutations, Fig. 3B, bottom) as it was observed in the hippocampus. Similarly comparison of responses to best familiar face compared with the corresponding voice in face-selective and voice-selective neurons were found to be biased toward the prime selectivity of the cell, as was observed in the hippocampus (Fig. 3C and Supplementary Fig. S11D). Overall faces and voices of same individuals are represented by distinct rather than same neurons in the monkey in the hippocampus and in TE.

**Conclusion 1**

At the single cell level, hippocampal activity to social stimuli appeared to be similar to the activity observed in the inferotemporal cortex. In both regions, we found: (1) the existence of neurons responding differently to social stimuli compared with nonsocial stimuli, but mainly in the visual rather than acoustic modality, (2) that responses to voices are less informative than responses to faces about the subcategory they code for, (3) that neuronal responses to facial and vocal identities are poorly correlated. However compared with inferotemporal neurons, hippocampal neuron tuning was broader, in the sense that they exhibited less fine tuning toward the different social categories. We next tested if these results translate at the population level, by analyzing the activity of all the recorded neurons with unsupervised analyses.

**Population Analysis**

For each of the 343 recorded neurons, the $z$-score for each stimulus was calculated as the mean neuronal response to the stimulus divided by its standard deviation. To determine which parameters are driving neuronal activity at the population level in both brain areas, principal component analyses were
performed. To this end, all the stimuli from the set were used as observations, and the $z$-score from each cell for each stimulus of the set were defined as the variables. Projection of stimulus identity on the first 2 principal components of neuronal population $z$-scores revealed that, visual and auditory stimuli are segregated by neuronal activity, in both regions (Fig. 5A and Supplementary Fig. S12). This segregation was confirmed by clustering with a Gaussian mixture model of the stimuli projection maps. When calculating the Euclidean distance in the principal component space between neuronal responses to every pair of stimuli and further linking the stimuli hierarchically into a cluster tree, we also found similar cluster trees for both regions, with distinction between visual and auditory stimuli (Fig. 5B). Dynamics of neuronal activities also presented dissimilarities across modalities in both regions, with more neurons reaching their peak of activity early in the stimulus presentation period for visual stimuli while peaks of activity were widely distributed along the auditory stimuli presentation (Fig. 5C), indicating that responses to acoustic stimuli were not time-locked to the stimulus onset.

Analysis of the stimuli projection maps (Fig. 5A and Supplementary Fig. S12) and hierarchical trees (Fig. 5B) revealed that not only auditory and visual stimuli but also face and object stimuli appear segregated by neuronal population activities in both regions. The optimal clustering of stimuli, assessed by minimizing Bayesian information of the Gaussian mixture model, was constituted of 3 components comprising acoustic stimuli (blue-green marks), facial stimuli (red-pink circles) and object stimuli (red-pink crosses, Fig. 5A). In hippocampus, using the same methodology, the best clustering was obtained along the modality dimension. However, note that points corresponding to objects (red-pink crosses) were not included in the lower visual cluster suggesting that all visual stimuli were not coded in the same manner and may be functionally distinguished into the categories of faces and objects (Fig. 5A). In contrast to visual stimuli, neuronal activities to voice and sound stimuli clustered together in each region (Fig. 5A), indicating that both regions map and segregate in-between visual stimuli rather than in-between auditory stimuli. It also suggests, that compared with face-selective cells, voice-selective cells did not drive enough population coding to distinguish between vocal and sound stimuli, possibly due to their less pronounced increase in FRs compared with their baseline activity (2-sided t-test, $P = 0.02$ in TE, $P = 0.028$ in the hippocampus, Fig. 5C). In the hierarchical trees (Fig. 5B), vocal stimuli also appear to cluster all-together (blue-green cluster), whereas visual stimuli appear to cluster in 3 groups: a first one encompassing most of the face stimuli (magenta cluster) and 2 encompassing mainly visual objects (yellow-orange and yellow-green clusters).

Conclusion 2

At the population level, hippocampal activity to social stimuli appears to be similar to the activity observed in the inferotemporal cortex. In both regions, we find that: (1) visual stimuli cluster separately from auditory stimuli, (2) they differ in their response dynamics, and (3) faces cluster apart from visual objects while voices and nonvocal sounds cluster together.

Discussion

Faces and voices are the major cues read out and used by primates (including humans and rhesus monkeys) to maneuver in their social environment. They provide essential social information about others’ status, such as identity, gender, species, familiarity, etc. (Chazanfar and Santos 2004; Belin 2006; Leopold and Rhodes 2010). In this study, we contribute to the goal of characterizing how hippocampal neurons in the monkey are activated when monkeys are exposed to these social stimuli through pictures and audio samples. At the single cell and population levels, hippocampal activity to social stimuli appeared to be much more similar to the activity observed in the monkey inferotemporal cortex than we expected. It differed from the identity selective activity observed in human epileptic patients, but it also differed from the poor sensitivity to social stimuli observed in rats. Thus in the monkey hippocampus, we found: (1) the existence of neurons responding differently to social stimuli compared...
Evidence for the Existence of Neurons Responding Differently to Social Stimuli Compared with Non-social Stimuli

In humans, the hippocampus might play a role in building and maintaining social relationships and networks (Allen and Fortin 2013), as evidenced by study of human patients with hippocampal impairment who present limited social circles compared with unimpaired controls (Davidson et al. 2012) and by single-unit recordings of human hippocampal cells coding for familiar and well-known individuals (Quiroga et al. 2005; Quian Quiroga et al. 2009; Viskontas et al. 2009). In rats, both lesion studies (Becker et al. 1999) and single-cell recording (von Heimendahl et al. 2012; Zynyuk et al. 2012) rather pointed to a nonimplication of the hippocampus in social representation. While in monkeys, lesions studies remained unclear (Machado and Bachevalier 2006), the present single-unit study reveals that monkeys hippocampus might play a role in social representation, because we show that hippocampal neurons expressed a significant enhancement of their activity to faces or voices, even though their processing was not relevant to the task. Although tested with an unequal number of items for each category, the distinction between faces and objects was sufficiently robust to also be attested at the level of the population when data were analyzed in an unsupervised manner. In this respect, monkey hippocampus appears to participate in representing social information and its role looks more similar to that of the human hippocampus than to that of the rat’s one. This result converges with observations made with another technique, functional magnetic resonance imaging (fMRI), showing the presence within the hippocampus of an area having an enhanced activity for faces compared with objects in monkeys (Ku et al. 2011; Lafer-Sousa and Conway 2013) and in humans (Ishai et al. 2005).
studies using faces and voices (Holdstock et al. 2010; Joassin et al. 2011; Love et al. 2011) and with single-cell recordings using faces and nouns (Quian Quiroga et al. 2009). We thus wondered if single-neuron correlates of cross-modal association of identity could be found in the monkey hippocampus. Both at the single-unit and at the population level, we found that neuronal activity to faces and voices differed significantly. Activity for faces was robust, distinct from that to objects and coding for subcategories of faces. On the contrary, activity for voices was characterized by lower increase in the FR and was found poorly specific. This might come from the fact that voices have been shown to be weaker cues than faces for identifying individuals in humans, both after having been learned in conjunction with faces in a multi-modal setting, as it happens in real-life, (Hanley et al. 1998; Hanley and Turner 2000; Damjanovic and Hanley 2007) and when learned separately from faces in a unimodal setting (Olsson et al. 1998; Joassin et al. 2004; von Kriegstein and Giraud 2006; Mullenix et al. 2009). Since hippocampal neuronal activity might represent memory recall triggered by sensory cues, this could explain why “known” voices elicited fewer responses than “unknown” faces in our study, as well as why “unknown” voices elicited fewer responses than “unknown” faces. However, differences in neural coding for voices compared with faces have also been observed in other studies investigating other brain regions. For example, voice-selective cells, recorded in the voice area found with fMRI in monkeys (Perrodin et al. 2011) or around it (Kikuchi et al. 2010), are less numerous than face-selective cells found in face areas, they also respond to a smaller amount of voice extracts than face-selective cells do to face exemplars (Baylis et al. 1985; Hasselmo et al. 1989; Rolls and Tovee 1995), and have weaker enhancement in the FR compared with face-selective cells recorded in TE. Thus, the less distinct responses to [vocalizations vs. object-sounds] compared with [faces vs. objects] that we observe in the hippocampus might also alternatively be a perpetuation of different inputs received by the hippocampus from these lower-level regions. Alternatively, the absence of an auditory evoked response can also suggest that the responses to auditory stimuli might arise as feedback signals rather than from feed-forward input. Finally, the behavioral task performed by the monkeys was different during visual and auditory stimuli presentations. While monkeys could inspect the images by freely moving their gaze, they were required to fixate during sound presentation. It could be argued that fixation would lower neuronal activity to the stimuli, as it has been observed in the superior colliculus (Bell et al. 2003), and on the contrary that free-viewing would enhance neuronal responses. However, the free-viewing responses to faces we observe in the hippocampus and the inferotemporal cortex are in the range of those seen with fixation in other studies in the inferotemporal cortex (Baylis et al. 1985; Hasselmo et al. 1989; Rolls and Tovee 1995); also the study having discovered the voice area housing most of the voice-selective neurons has been carried while monkeys were required to fixate (Petkov et al. 2008).

No Evidence of a Cross-modal Coding in the Monkey Hippocampus

While visual and auditory responses appeared being coded at different scales, they could still potentially match qualitatively. By normalizing for mean response amplitudes in each modality, we examined if neuronal responses to facial identities would have a correspondence in the neuronal responses to vocal identities. We found that activities for facial and vocal identities were poorly correlated. In this regard, our results differ from the ones described in human studies, in which same neurons coded for facial identity and for written or orally pronounced noun. It is possible that the higher activity for visual stimuli, compared with acoustic ones, arises from the location of our recording sites within the hippocampus in its more anterior part, which is densely and directly connected to higher visual areas (Rockland and Van Hoesen 1999; Yukie 2000; Zhong and Rockland 2004) and only more sparsely and mainly indirectly to higher auditory areas through polymodal areas (Zhong et al. 2005; Mohedano-Moriano et al. 2007, 2008; Munoz-Lopez et al. 2010).

Auditory stimuli might be preferentially coded in a more posterior part of the hippocampus (Gil-da-Costa et al. 2004; Peters, Koch et al. 2007; Peters, Suchan et al. 2007) which we lacked sampling of. Nevertheless, the cross-modal cells found in humans were sampled throughout the medial temporal lobe including both anterior and posterior parts of the human hippocampus (Quiroga et al. 2005; Quian Quiroga et al. 2009). The specifically human aptitude for naming persons and things, has been shown to participate in the way humans represent, memorize and retrieve complex associations such as full autobiographical episodes (Clark and Squire 2013). It could be wondered if this aptitude also translates in the way cross-modal associations are represented in higher-level areas such as the hippocampus and further, if monkeys and others animals who rely on unnamed associations might lack an explicit coding at the single-cell level for cross-modal identity association. If we missed an explicit code by not prescreening the neurons but if this code yet exists in neurons we have not recorded from, it is not observable at the population level.

Less Sparse Coding of Identity Than in Humans

Because visual and auditory modalities appeared to be coded separately, we investigated invariant representation within modalities, throughout viewpoint or audio extracts. Neurons signaling facial identity were observed as evidenced by view-invariant coding. However, most neurons responded to more than one individual. This result is in accordance with early unsupervised recordings from human hippocampal neurons showing nonsparse hippocampal categorization of visual stimuli including faces (Fried et al. 1997, 2002; Kreiman et al. 2000). Similarly to these studies, we observe that 40% of hippocampal cells respond to stimuli and exhibit a medium range of selectivity (0.5 or 0.6). More recent findings by the same group (Quiroga et al. 2005; Waydo et al. 2006; Quian Quiroga et al. 2009) showed that, when using a supervised recording protocol precisely searching for this type of cells, few ultra-sparse cells (10%) in the human hippocampus represent single individuals in a selective and invariant manner. Because we have not used this type of supervised recording protocol, it is possible that we missed conceptual coding cells, for example, activated in the same manner only by all the pictures of a well-known monkey’s face, or only by the audio extracts of a monkey’s voice, or a fortiori only by all the pictures of a well-known monkey’s face and all the audio extracts of the same monkey’s voice.

A Representation of Social Stimuli Which Resembles That of the Inferotemporal Cortex

Recordings of inferotemporal cells, with the exact same stimulus set, material, animals, and protocol conditions, revealed that stimuli were coded in the monkey hippocampus in a way much more similar to the representation already known in the inferotemporal cortex than in the human hippocampus. First, we
A Categorization of Facial Stimuli Which Resembles That of the Inferotemporal Cortex

Hippocampal face-selective cells tuning was broader than that of inferotemporal face-selective cells, though a proportion of them exhibited a fine tuning toward facial categories, mainly that of species, familiarity, and viewpoint. Modulation by these characteristics has already been observed in monkey inferotemporal cortex, a result which we reproduce here (species: Sigala et al. 2011; viewpoint: Perrett et al. 1985; De Souza et al. 2005; Freiwald and Tsao 2010, familiarity: Eifuku et al. 2011). Hippocampal neurons in monkeys therefore appear to categorize visual stimuli into a space encompassing perceptual components (face-object, species, or viewpoint distinctions) as well as cognitive components (familiarity). However, this observation is not specific to the hippocampus since familiarity is found to also modulate activity of a portion of face-selective cells in TE, which is in line with fMRI results in TE (Sugiura et al. 2001; Rotshtein et al. 2005) and the hippocampus (Rotshtein et al. 2005; Denkova et al. 2006). In other words in our protocol, this part of the hippocampus does not appear to be more peculiar for mnemonic processes than the inferotemporal cortex. On the contrary, facial information probably reached the hippocampus through feed-forward projections from the temporal cortex. In the hippocampus, the neuronal response to faces was globally less selective, but maintained important information about species or familiarity. Alternatively, hippocampal response specifics might also come as feedback from the amygdala, which is anatomically close and also represent facial information (Rolls 1984; Leonard et al. 1985; Gotthard et al. 2007; Hoffman et al. 2007; Mosher et al. 2010; Rutishauser et al. 2011; Hadj-Bouziane et al. 2012). However, compared with the monkey and human amygdala, where it was found that a majority of facial responses were inhibitory (Rutishauser et al. 2011) and that the “depth” of an inhibitory response can be just as informationally rich as the height of an excitatory response (Mosher et al. 2010), we have neither found these proportions in the monkey hippocampus (or area TE), nor these specificities.

Conclusion

Codes for social stimuli appear to be present throughout the temporal lobe until the hippocampus: faces and voices seemed to have colonized some processing resources of the monkey hippocampus. However, in an anterior part of the primate hippocampus, cognitive maps of visual—rather than auditory or bimodal—stimuli predominate, probably participating in the read-out of individuals’ facial information rather than vocal one. An explicit conceptual coding of identity, as discovered in human hippocampus, has not been observed in the present study. These conclusions were not available in the preexistent literature, which focused mainly on secondary visual/auditory regions of the temporal cortex for face and voice processing while monkey hippocampal neurons were studied for their involvement in memory or navigation. The hippocampus plays a crucial role in autobiographical/episodic memory and social events make up those memories to a large extent. Thus the presence of cells representing social cues in the hippocampus constitutes the prime step to understand how social cues are incorporated into memories. The existence of neurons explicitly coding identity throughout modalities might be found in more posterior parts of the hippocampus or in other parts of the monkey brain (Kuraoka and Nakamura 2007; Romanski 2012).

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

Funding

This work was supported by a Marie Curie reintegration grant and a salary grant from Fondation pour la Recherche Médicale to S.W.; a PhD grant cofinanced by Centre National de la Recherche Scientifique and by Direction Générale de l’Armement to J.S.; Fondation pour la Recherche Médicale; Association des Femmes Françaises Diplômées d’Université-Dorothy Leet; Fondation Bettencourt-Schuller to J.S.; and a grant by Agence Nationale de la Recherche BLAN-1431-01 and ANR-11-IDEX-0007 to J.R.D.

Notes

The authors thank S. Wiener for helpful comments on earlier versions of the manuscript; P. Baraduc for discussion on statistical analyses; J.-L. Charieau and F. Hérant for animal care. Conflict of Interest: None declared.