NeuroImage 125 (2016) 280–290

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Fast ventral stream neural activity enables rapid visual categorization

Maxime Cauchoix [a,b,c,*,1], Sébastien M. Crouzet [b,c,d,1], Denis Fize [b,c], Thomas Serre [d]

[a] Institute for Advanced Studies in Toulouse, France
[b] Centre de Recherche Cerveau et Cognition, Université Paul Sabatier, Université de Toulouse, Toulouse, France
[c] Faculté de Médecine de Purpan, CNRS, UMR 5549, Toulouse, France
[d] Cognitive, Linguistic and Psychological Sciences Department, Institute for Brain Sciences, Brown University, USA

ARTICLE INFO

Article history:
Received 23 April 2015
Accepted 7 October 2015
Available online 20 October 2015

Keywords:
Rapid categorization
Object recognition
Ventral stream
Monkey electrophysiology
Natural scenes

ABSTRACT

Primates can recognize objects embedded in complex natural scenes in a glimpse. Rapid categorization paradigms have been extensively used to study our core perceptual abilities when the visual system is forced to operate under strong time constraints. However, the neural underpinning of rapid categorization remains to be understood, and the incredible speed of sight has yet to be reconciled with modern ventral stream cortical theories of object recognition.

Here we recorded multichannel subdural electrocorticogram (ECoG) signals from intermediate areas (V4/PIT) of the ventral stream of the visual cortex while monkeys were actively engaged in a rapid animal/non-animal categorization task. A traditional event-related potential (ERP) analysis revealed short visual latencies (<50–70 ms) followed by a rapidly developing visual selectivity (within ~20–30 ms) for most electrodes. A multi-variate pattern analysis (MVPA) technique further confirmed that reliable animal/non-animal category information was possible from this initial ventral stream neural activity (within ~90–100 ms). Furthermore, this early category-selective neural activity was (a) unaffected by the presentation of a backward (pattern) mask, (b) generalized to novel (unfamiliar) stimuli and (c) co-varied with behavioral responses (both accuracy and reaction times). Despite the strong prevalence of task-related information on the neural signal, task-irrelevant visual information could still be decoded independently of monkey behavior. Monkey behavioral responses were also found to correlate significantly with human behavioral responses for the same set of stimuli.

Together, the present study establishes that rapid ventral stream neural activity induces a visually selective signal subsequently used to drive rapid visual categorization and that this visual strategy may be shared between human and non-human primates.

Introduction

The robust and accurate categorization of natural object categories is critical to survival, as it allows an animal to generalize many properties of an object from its category membership (Fabre-Thorpe, 2003; Fize et al., 2011; Rosch, 1975; Thompson and Oden, 2000; Zentall et al., 2008). Human and non-human primates excel at visual categorization: They can rapidly and reliably categorize objects embedded in complex natural visual scenes in a glimpse (see Fabre-Thorpe, 2011; Potter, 2012 for recent reviews).

It is well known that object recognition is possible for complex natural scenes viewed in rapid visual presentations that do not allow sufficient time for eye movements or shifts of attention (Biederman, 1972; Potter and Levy, 1969; Thorpe et al., 1996). The underlying visual representation remains relatively coarse as participants frequently fail to localize targets that are correctly detected in an image stream (Evans and Treisman, 2005). Studies using backward-masking (Bacon-Macé et al., 2005) and saccadic responses (Crouzet et al., 2010; Kirchner et al., 2009) have further demonstrated that recognition is possible under severe time constraints. While much is known about the psychological basis of rapid categorization, much less is known about the underlying neural processes and, in particular, the timing of the corresponding perceptual decisions.

Using human scalp electroencephalography (EEG), Thorpe et al. (1996) first demonstrated that a category-selective signal can be isolated from frontal electrodes shortly after a stimulus is flashed (within ~150 ms post stimulus onset). Previous work using intra-cranial recordings has shown that it is possible to decode object category information from the ventral stream of the visual cortex very rapidly (within ~100 ms post stimulus onset) in both humans (Liu et al., 2009) and monkeys (Hung et al., 2005; Kreiman et al., 2006; Vogels, 1999a). However, this work either used a passive viewing paradigm (Hung et al., 2005; Kreiman et al., 2006; Liu et al., 2009) or involved a relatively simple basic-level categorization task, such as trees vs. objects (Vogels, 1999a), and did not establish any link between ventral stream neural

* Corresponding author at: Institute for Advanced Studies in Toulouse, 21 allée de Brienne, 31015 Toulouse Cedex 6, France.
E-mail address: maxime.cauchoix@iast.fr (M. Cauchoix).
[1] These authors contributed equally to the work.

http://dx.doi.org/10.1016/j.neuroimage.2015.10.012
1053-8119/

activity and (speeded) behavioral responses during rapid categorization. Indeed, previous monkey electrophysiology work has found little co-variation between reaction time and neural latencies in the inferotemporal cortex (DiCarlo and Maunsell, 2005; Eifuku et al., 2004; but see also Mruczek and Sheinberg, 2007).

Here, we recorded ECoG activity in monkeys actively engaged in a rapid natural scene categorization task with the aim to characterize the time-course of visual processing and establish a link between fast ventral stream neural activity and rapid behavioral responses.

Material and methods

Successful learning of the categorization task

Protocol

Two male rhesus macaques (M1 and M2, both aged 14) performed the experiment. The animals were restrained in a primate chair (Crist Instruments, GA), sitting 30 cm away from a 1024 × 768 tactile screen. Stimuli (Fig. 1a) were flashed centrally for 33 ms covering about 7° of visual angle on a black background, with a 1.5–3 s random inter-trial time interval between successive images.

In masked trials (presented in separate blocks), a pattern image corresponding to visual pink noise was presented 50 ms post stimulus onset for 33 ms (Fig. 1b). Brief presentations, in addition to masking on specific trials, prevented exploratory eye movements and constrained the time available for information uptake.

The two monkeys performed a natural scene rapid categorization task by releasing a button and touching the screen when they saw an animal in the stimulus presented (target) or keeping their hand on the button otherwise (distractor). A drop of fruit juice rewarded correct responses; on incorrect trials, the stimulus was re-displayed for 3 s, thus delaying the next reward and motivating the animal to answer as fast and accurately as possible. All procedures conformed to French and European standards concerning the use of experimental animals. Protocols were approved by the regional ethical committee.

Animal training

Initial training followed the procedure previously reported in Fabre-Thorpe et al. (1998): learning was gradual, starting with 10 images and progressively introducing new scenes everyday over a period of several weeks until both monkeys were performing well on the familiar set of stimuli. While the monkeys' motivation and level of reward were kept stable by randomly interleaving familiar and new stimuli, the recurrent introduction of new stimuli (usually 10–20%) forced the animals to learn to generalize to novel stimuli rather than to rely on stimulus memorization. Both monkeys were trained for intermittent periods on the animal/non-animal task since 2005 (Fabre-Thorpe et al., 1998; Fize et al., 2011; Macé et al., 2005).

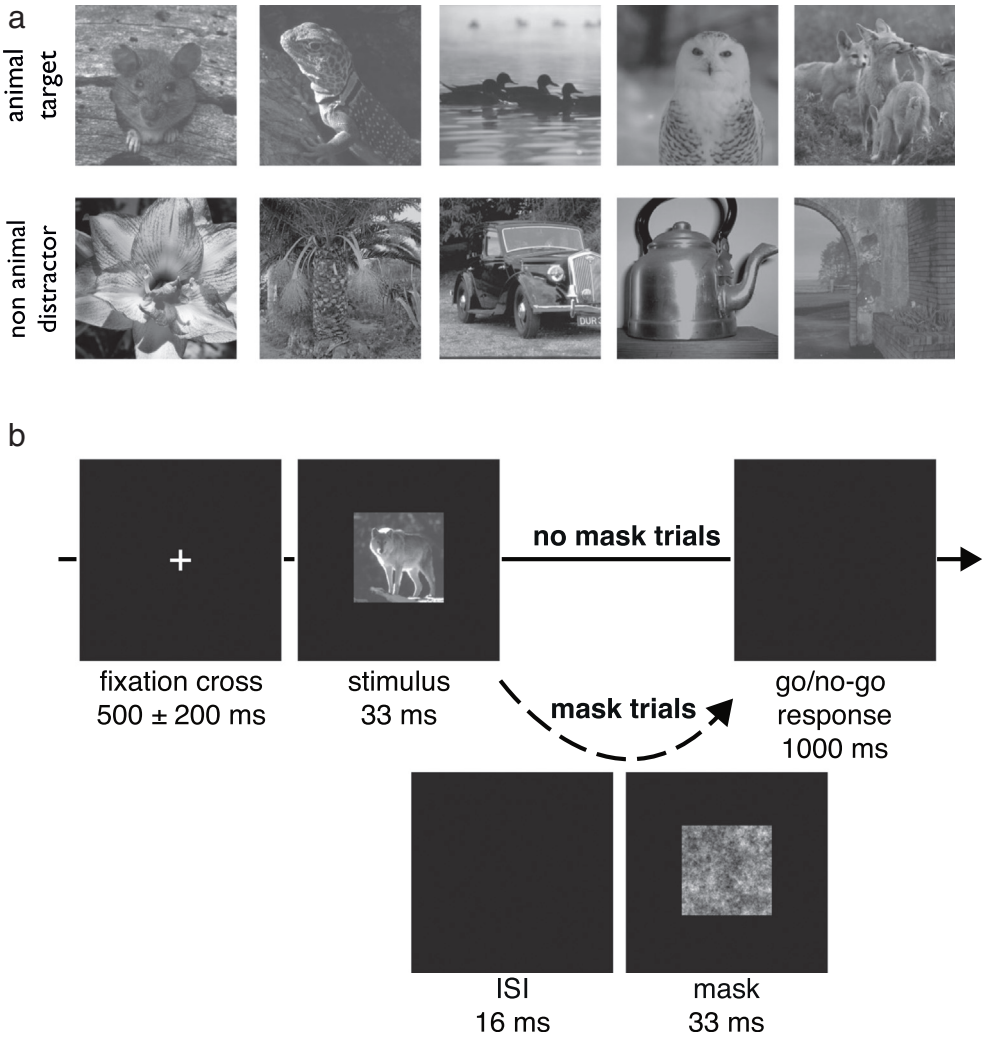

Fig. 1. Stimulus set and visual categorization task. a) The stimulus set consisted of natural gray-scale images with both animal targets (n = 340) and non-animal distractors (n = 340). b) Following the presentation of a fixation cross, an image was flashed for 33 ms. The monkeys indicated whether a target was present by releasing a button within 1 s following image presentation. On half the trials, a backward mask (1/f pink noise) was displayed for an additional 33 ms following a 16-ms blank screen (50 ms SOA).

To insure that the monkeys properly learned the concept of animal, diverse stimuli were sampled from different animal families including mammals (57%), birds (23%) insects (2%), reptiles and amphibians (8%) as well as fishes and crustaceans (10%). Exemplars were chosen to be as varied and perceptually dissimilar as possible.

Familiar stimulus set

The image set consisted of natural gray-scale photographs (256 × 256 pixels) equalized for average luminance and global contrast (root mean square over pixel intensities). A familiar stimulus set consisting of 280 images (140 animal targets and 140 non-animal distractors) was used to train the animals. An effort was made for target and distractor scenes to be as varied as possible. Targets included pictures of mammals, birds, insects, reptiles, or amphibians presented either in isolation or in groups and at various positions in the images. Distractors included pictures of various objects (trees, flowers, tools, etc.) in natural and man-made environments (mountain, sea, city, etc.).

Behavioral data analysis

We report behavioral accuracies as d′ (d′(t) = z[Hits(t)] − z[False Alarms(t)], where z[] is the inverse of the normal distribution function). To test whether the recognition accuracy of the two monkeys was above chance on images from the novel set, we used a χ^2 test against the null hypothesis that the numbers of correct and incorrect responses are equal. To compare the behavioral accuracy of the monkeys mask vs. no-mask trials, Pearson's χ^2 tests with Yates' continuity correction were used for each monkey separately.

We conducted an analysis of the time course of accuracy (d′) against reaction times. Hits and false alarms were binned independently (30 ms bins) according to reaction times. We then computed d′ scores for individual bins of these cumulative histograms. Minimal reaction times were computed using the first 30 ms time bin followed by at least five consecutive bins for which the number of hits significantly outnumbered the number of false alarms (binomial test; $p < 0.01$) as done previously in Liu et al. (2009). The timing of the effect of the backward mask was estimated using a Pearson's χ^2 test with Yates' continuity correction on individual bins. All analyses were performed using the R software (http://www.r-project.org).

Event-related potentials (ERPs) and visual receptive fields

ECoG recording and preprocessing

Monkeys were implanted with subdural macro-electrodes. Holes (1-mm diameter) were drilled in the animal skull under anesthesia, and electrodes were lowered manually based on initial anatomical MRI scans. The two monkeys where implanted in both hemispheres (9 and 13 electrodes for M1 and M2, respectively). A more precise electrode location was then subsequently assessed by merging X-rays (with both the electrodes and the skull visible) and T1-weighted MRI anatomical scans (see Fig. 2a and Fig. S1 for electrode locations).

The macro-electrodes used were steel wires with 150 μ of diameter and less than 1 MΩ impedance. All electrodes were connected to a DB plug. The entire system was attached to the monkey's skull using screws and dental cement. Recordings were performed using the NeuroScan EEG and SynAmps amplifier systems with a sampling rate of 1000 Hz (band pass 0.1–200 Hz). Frontal electrodes were used as reference (intra-cortical for M1, within the frontal sinus for M2). M1 was recorded head-free; M2 was head-restrained. Recording sessions spanned five consecutive days (3–4 blocks of 640 trials daily until water satiation), resulting in the collection of 7499 and 8757 trials for M1 and M2, respectively.

We used the EEGLAB toolbox (Delorme and Makeig, 2004) to import neural data in MATLAB in order to preprocess the recordings. A notch filter was used to remove 50 Hz noise. Trials used for subsequent analyses were selected based on both raw potentials (<3 standard deviations from the mean) and global ERP power (sum of squares computed over all electrode potentials for every time point); ~10% of the trials were discarded based on weak visual power between 60 and 120 ms post stimulus onset. The signal was baseline corrected [−50; 30 ms] trial by trial and down-sampled to 512 Hz.

ERP visual latency

Following the procedure described in Yoshor et al. (2007), we computed confidence intervals (99% CIs) for each electrode using the average activity computed during baseline (−50–30 ms time window; see Fig. 2b and Fig. S2). Visual latencies for individual electrodes were then determined by considering the first 2-ms time bin for which the average voltage response fell outside the 99% CIs.

Coarse receptive field mapping

We used the visual mapping procedure described in Yoshor et al. (2007) to map out the visual receptive fields in one animal (M1). M2 accidentally lost his electrode cap before we were able to conduct the mapping experiment. Small visual mapping stimuli (2° of visual angle white squares) were briefly flashed at different positions (10 × 10 grid covering an area of 20° × 20° of visual angle around the fixation cross) on screen at the beginning of the trial (during the fixation cross interval) while the monkey was waiting for the trial to start. On each trial, 3 to 7 mapping stimuli were presented (location sampled at random) for 16 ms with an inter-stimulus interval of 130 ms.

After baseline correction (−50–30 ms), we computed ERPs for each electrode and each location of the mapping stimuli using 15 trials per location. We used the maximal ERP response during the 30–130 ms stimulus interval as a measure of visual responsiveness for the corresponding stimulus location. Finally, we performed a two-dimensional Gaussian fitting procedure on the corresponding spatial maps for each electrode. RF width was determined by averaging the full width at half height for each of the 2 axes from the fitted Gaussian. The receptive fields of all but one V4 electrode could be successfully mapped using this procedure and a criteria for goodness-of-fit described in (Yoshor et al., 2007) (Fig. S3). The fact that we were able to isolate spatially restricted receptive fields in hemifields consistent with the hemisphere location of the corresponding electrodes (Fig. S3) suggests that the procedure was indeed successful.

Fast single-trial decoding of superordinate category

Multi-variate pattern analysis (MVPA) was performed directly on the neural signal in two ways. A global decoding accuracy measure was obtained by feeding a classifier with the entire 90–140 ms IFP waveforms from all electrodes. A temporal decoding accuracy measure was obtained by feeding each time bin of the IFP signal from all electrodes. Such temporal decoding characterizes the temporal evolution of the category signal.

For both neural decoding and the computational model, a linear Support Vector Machine (SVM) classifier was used. The classification procedure ran as follows: (1) The image set was equally split into a training set and a test set that each contained an equal proportion of target and distractor images. (2) An optimal cost parameter C was determined through line search optimization using 8-fold cross-validation on the training set of images. (3) An SVM classifier was trained and tested on each split.

The reported results correspond to the average performance (non-parametric 95% confidence intervals of the mean) using a cross-validation procedure (n = 300) whereby different training and test sets were selected each time at random. A measure of chance level was obtained by performing the same analysis on permuted labels. The decoding was considered above chance when 95% of the differences between decoding on true or permuted labels paired over the 300 cross-validations fell above zero. When comparing different conditions, such as masked and unmasked trials, significant differences were assessed using a non-parametric permutation test at each time-point (for those

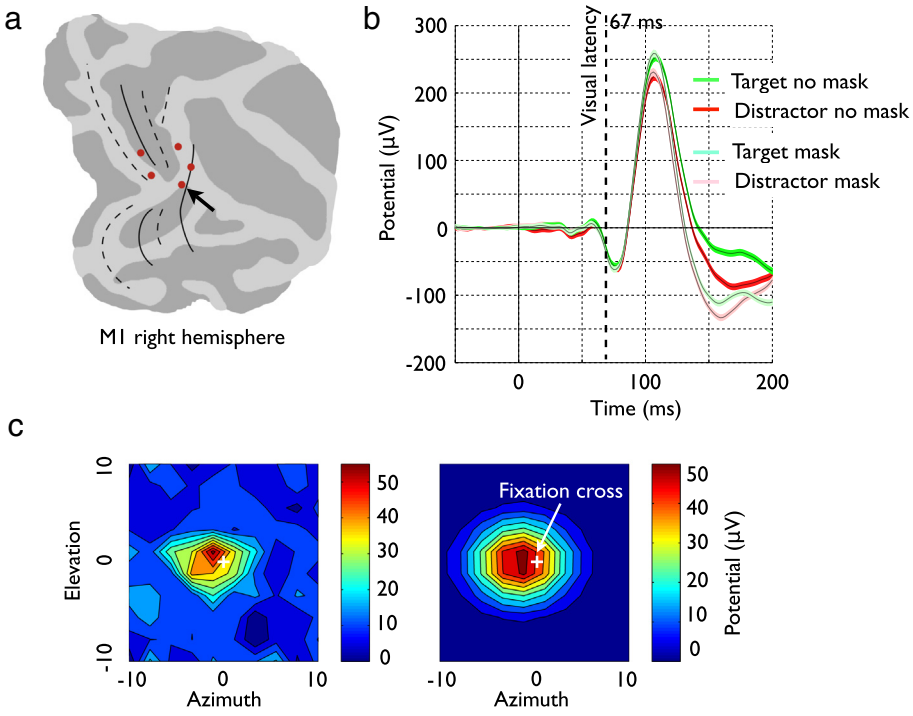

Fig. 2. Sample electrode locations, event related potential and visual receptive field. a) Electrode tips (red dots) are shown on (FreeSurfer) flat maps for monkey M1 (right hemisphere). b) The corresponding ERP for the electrode marked with an arrow in panel a. Shown is the average potential evoked by animal targets (green) vs. non-animal distractors (red) for both the mask (darker shade) vs. no-mask (lighter shade) conditions. Thin black lines indicate sample averages with corresponding CIs (95%) obtained via bootstrapping shown with transparency. c) Visual receptive field for the same electrode before (left) and after (right) two-dimensional Gaussian fitting.

time points that exhibited a significant baseline decoding). In the case of temporal decoding, to correct for multiple comparisons, a bin was considered significant when followed by at least five consecutive significant bins ($p < 0.01$), as done in Liu et al. (2009).

Backward masking: comparison between neural decoding and behavior

The protocol used here followed the description given in Protocol section and was adapted from a human psychophysics experiment (Serre et al., 2007). During masked trials (presented in separate blocks), a pattern image corresponding to 1/f random noise mask was presented 50 ms post stimulus onset for 33 ms. The mask was generated by filtering random noise through a Gaussian filter. The order of presentation of mask and no mask blocks was alternated and counterbalanced between days.

Fast ventral stream neural activity enables generalization to novel exemplars

To test the ability of animals and neural decoding to generalize to unfamiliar stimuli, novel images never seen before by the monkeys were introduced to test their ability to generalize to new stimuli. This set contained 400 images total sampled from a larger set used in (Serre et al., 2007) divided equally between animal target and non-animal distractors. This set of images was introduced to the monkeys gradually during recording sessions on five consecutive days. Each day, the animals were tested using 1/4 images sampled from this novel set (80 stimuli selected each day at random without replacement) and 3/4 familiar images (240 stimuli selected at random from the familiar set). This set of randomly mixed familiar and novel images was repeated twice in a row to form a 640-trial block. Each day, monkeys performed 3 to 4 of those blocks until water satiation.

Robustness of the neural decoding to background clutter

The novel (unfamiliar) set of images was further subdivided into four equal-sized image subsets corresponding to different viewing distances of the animal (or distractor object): “head,” “close-body,” “medium-body,” and “far-body”. A set of distractors with matching mean distance from the camera was selected from a database of annotated mean-depth images (see Serre et al., 2007 for details). Because the size of the images is fixed, the animal viewing distance provides a good proxy for the amount of background clutter present in these images.

Fast single-trial decoding is predictive of behavioral responses

The animalness scores reflect how likely each individual image in the stimulus set is to be classified as an animal irrespective of its actual category label. Here we used this score to compare classification based on neural, behavioral and model data (Fig. S4). An animal score was computed from behavioral data by considering the fraction of trials for which this specific image was classified as (animal) target by the monkeys. For both neural and model data, an animalness score was computed by considering the decision value of the linear (SVM) classifier used for MVPA. The output of the classifier for a specific image corresponds to the distance between this image and the separating hyperplane and can be used as a direct measure of accuracy. Classifier outputs were thus computed for each image of the test set over 50 cross-validations, and then averaged to obtain one single animalness score per image. Animalness scores across all stimuli for behavioral, neural and model data were directly correlated using both Spearman correlation (reported as r^2) and partial correlation (r^{2*}) measures (using MATLAB built-in functions) to control for semantic/category effects. Such partial correlation measures reflect the intrinsic correlation between, for example, the neural and model data, due to visual information beyond predicted

category labels. Using this correlation measure, it is thus possible for two visual systems to exhibit a similar level of accuracy and be uncorrelated.

Fine visual information can be decoded independently of monkey behavior

We considered four (basic) categories of images from a subset of the target stimuli presented (people, macaques, chimpanzees, otter faces). Each subset contained 10 images for a total of 40 images. We trained and tested a classifier to discriminate between these four subcategories of animal images using a one-versus-all decoding procedure.

Results

Successful learning of the categorization task

Two monkeys (M1 and M2) were trained to perform a rapid categorization task in which they had to report the presence or absence of an animal in briefly presented natural scenes (see Material and methods; Fig. 1). Despite the large variability of the natural stimuli used, both monkeys were able to learn to categorize images with a high degree of accuracy (measured as d′; M1: 3.15; M2: 2.96) and very short median reaction times (RTs; M1: 305 ms; M2: 289 ms).

Event-related potentials (ERPs) and visual receptive fields

While the two animals were engaged in the rapid categorization task, we recorded subdural electrocorticogram (ECoG) signals from multiple electrodes implanted over intermediate areas (V4/PIT) of the visual cortex (Figs. 2 and S1). We estimated the visual latency of individual electrodes using the method described in Yoshor et al. (2007). Most electrodes exhibited short visual latencies ranging between ~40–70 ms post stimulus onset for M1 (median: 53 ms) and from ~35–80 ms for M2 (median: 59 ms, with a trend for V4 electrodes to exhibit shorter latencies than PIT electrodes) (Fig. S2).

To verify that the recorded neural responses were visual in nature, we mapped out receptive fields (RFs) coarsely in one of the two animals (M1) by flashing small white squares during the fixation-cross (pre-trial) intervals (see Material and methods). RF sizes ranged between 3.4°–8.9° with an average of 7.0° and were of similar size for V4 and PIT electrodes (Fig. S3, see Fig. 2c for an example electrode).

We further estimated the earliest significant differential activity for animal vs. non-animal stimuli for each individual electrode using a point-by-point analysis as done in Liu et al. (2009) and Thorpe et al. (1996): Latency here is defined as the first time point where five consecutive points (10 ms bins) yields significance ($p < 0.01$) in a t-test. For one of the two monkeys (M1), all electrodes exhibited a significant differential activity for animal vs. non-animal under 200 ms. For the other monkey (M2), only about half of the electrodes (8 out of 13) exhibited significance. The earliest significant animal/non-animal differential activity found using this method occurred at 83 ms in M1 on one V4 electrode and at 89 ms in M2 for a PIT electrode (Fig. S2).

Fast single-trial decoding of superordinate category

The analysis above required to average each individual electrode neural signal across multiple trials. In some cases, such averaging may lead to data distortions that can be avoided using multi-variate pattern analysis (MVPA) combined with modern statistical methods (Meyers and Kreiman, 2012; Rousselet and Pernet, 2012). MVPA typically provides higher statistical power over uni-variate methods by pooling information across all electrodes, enabling reliable estimates of latencies from single trials (Cauchoix et al., 2012).

Here, we trained and tested a linear Support Vector Machine (SVM) classifier directly on pooled electrode potentials for every time point independently (Chang and Lin, 2011). A similar analysis was previously used in combination with human MEG (Cichy et al., 2014; Isik et al., 2014), EEG (Cauchoix et al., 2014) and ECoG (Liu et al., 2009) data as well as monkey electrophysiology data (Hung et al., 2005) and provides a compact unbiased summary of the entire time-course of visual processing (Fig. 3a). Reliable decoding of superordinate category information (animal vs. non animal) was possible from single trials under 100 ms post stimulus onset (M1: 92 ms, M2: 96 ms). Furthermore, electrodes from both V4 and PIT contributed to the decoding accuracy (Fig. 4).

Next, we tested the robustness of this early category-selective neural to the presentation of a backward (pattern) mask as used in several previous rapid categorization psychophysics experiments (Bacon-Macé et al., 2005; Serre et al., 2007; VanRullen and Koch, 2003).

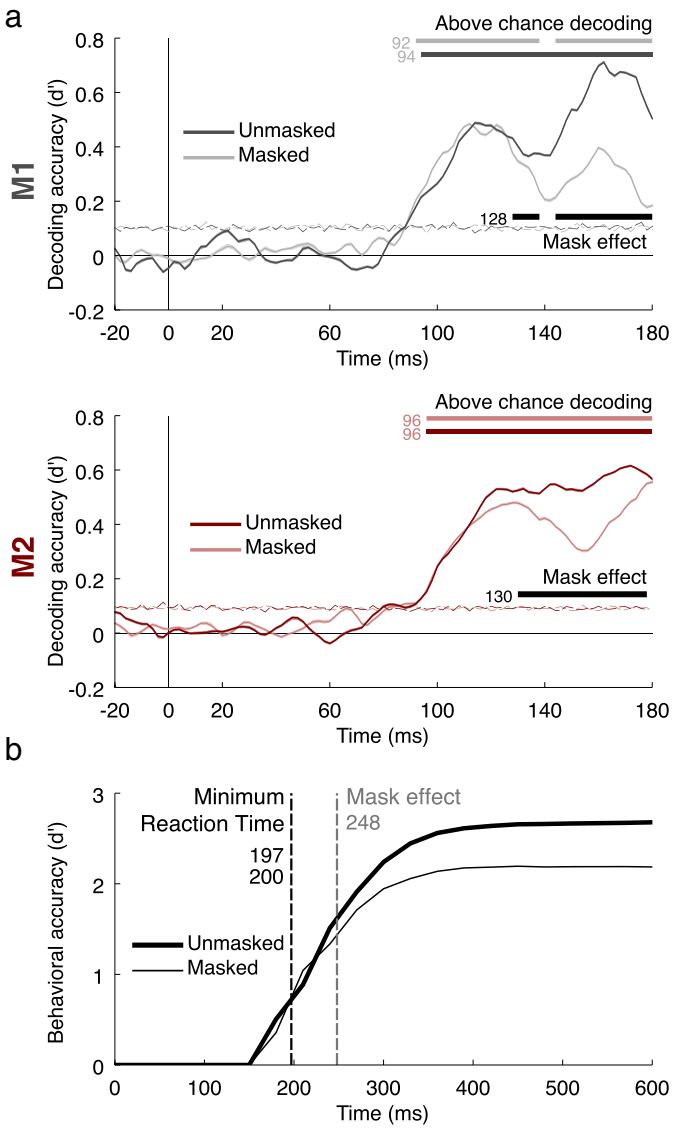

Fig. 3. Decoding from ventral stream neural activity and backward mask effect. a) Readout accuracy for monkey M1 (gray) and M2 (red) during the mask (lighter shade) vs. no-mask conditions (darker shade). Center lines correspond to the average decoding accuracy estimated across multiple cross-validations of the data. Corresponding CIs (95%) obtained via bootstrapping are shown with transparency (most confidence intervals are actually too small to be visible). Horizontal lines indicate decoding latencies for the mask and no-mask conditions (upper colored bars) as well as the earliest significant effect associated with the presentation of the mask on decoding accuracy (lower black bar on each plot). Horizontal dotted lines indicate the upper limit of the 95% CI around chance level obtained with a permutation procedure. b) Cumulative d′ curves plotted as a function of response times for mask (thin) vs. no-mask conditions (thick). Black dotted vertical line indicates minimum reaction times for mask (thin) vs. no-mask conditions (thick). Gray vertical line indicates earliest mask effect.

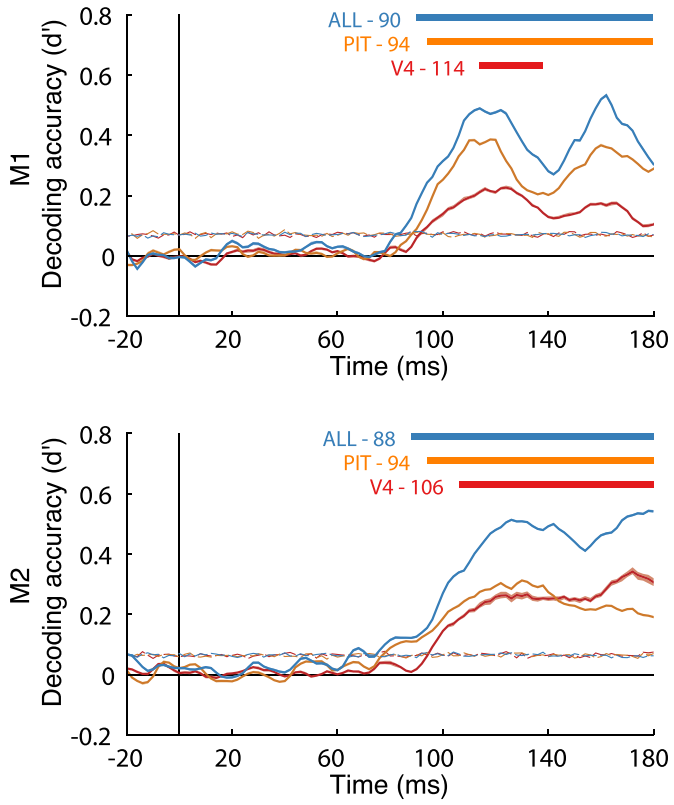

Fig. 4. Decoding comparison between V4 and PIT. Readout accuracy for V4 (red) vs. PIT (orange) and comparison with readout from all electrodes (blue) for monkey M1 (upper graph) and M2 (lower graph). Center curves correspond to the average decoding accuracy estimated over multiple cross-validations of the data. Corresponding CIs (95%) obtained via bootstrapping are shown with transparency (most confidence intervals are actually too small to be visible). Significant deviation from chance level is shown at the top of the graph with horizontal bars along with the corresponding earliest latency. Horizontal dotted curves indicate upper limit of the 95% CI around chance level obtained with a permutation procedure. This analysis was based on 9 electrodes for M1 (6 in V4 and 3 in PIT), and 13 for M2 (10 in V4 and 3 in PIT). The classification was nonetheless based on the same number of electrodes (n = 3) for V4 and PIT by using only 3 random electrodes from V4 (different samples used for different cross-validations of the data).

Backward masking: comparison between neural decoding and behavior

On half of the trials, a backward mask (1/f pink noise) was presented following stimulus presentation with a stimulus onset asynchrony (SOA) of 50 ms. This type of mask puts very severe time constraints on the visual system: It is assumed to interrupt visual processing by disrupting visual persistence and hindering recurrent signals (Keysers et al., 2001; Lamme and Roelfsema, 2000), thus encouraging fast, i.e. mostly feedforward, behavioral responses (VanRullen and Koch, 2003). Consistent with this hypothesis, previous human psychophysics studies have established that backward masking with an SOA ~50 ms (as used here) or longer has very little effect on the fastest human behavioral responses (VanRullen and Koch, 2003) and early scalp electro-encephalography (target vs. distractor) differential activity (Bacon-Macé et al., 2005).

Here, the overall behavioral accuracy (d′) of the two animals (computed on both familiar and novel images) was significantly reduced on masked trials (M1: 2.12 vs. 2.83 for mask vs. no-mask presentations, $\chi^2(1) = 99$, $p < 0.001$; M2: 2.14 vs. 2.55 for mask vs. no-mask presentations, $\chi^2(1) = 40$, $p < 0.001$) but far exceeded chance level (M1: $\chi^2(1) = 2059$, $p < 0.001$; M2: $\chi^2(1) = 2326$, $p < 0.001$).

Cumulative d′ curves plotted as a function of response times are shown in Fig. 3b. The cumulative number of hit and false alarm responses was used to calculate an accuracy measure at each time point t. Such analyzes of the time course of reaction times aim to provide a behavioral characterization of the underlying processing dynamics. The shortest RT was unaffected by the presentation of the mask (min RT = 197/200 ms with/without mask). A significant difference between masked and unmasked trials only appeared for trials with slower responses (with a delay roughly equal to the SOA ~50 ms; Fig. 3b).

An ECoG decoding analysis performed separately on trials with/without masking revealed very similar estimates of decoding latencies (M1: 94 ms, M2: 96 ms). Interestingly, the first significant difference between the two conditions was found during a later time window (starting at 128 ms post stimulus onset for M1 and 130 ms for M2; Fig. 3a). The corresponding ~30 ms initial neural processing time window left unaffected by the presentation of the mask matches well with the ~50 ms initial response time window estimated from behavioral responses above.

Next, we assessed how behavioral and neural responses generalize to novel (unfamiliar) stimuli, a hallmark of high-level abstract category formation beyond rote learning.

Fast ventral stream neural activity enables generalization to novel exemplars

We then evaluated behavioral responses and neural decoding accuracy separately for a novel and familiar set of images. On the very first presentation of the novel stimuli (80 images × 5 days), the accuracy of the monkeys (d′) remained well above chance (M1: 1.62, $\chi^2(1) = 96$, $p < 0.001$; M2: 1.99, $\chi^2(1) = 123$, $p < 0.001$). Similarly, global decoding from early ventral stream neural activity (90–140 ms time window post-stimulus onset) generalized well above chance from the familiar (M1: 0.68, $p < 0.01$; M2: 0.79, $p < 0.01$) to the novel set (M1: 0.34, $p < 0.05$; M2: 0.48, $p < 0.05$).

These results suggest that the two monkeys did form a high-level concept of the animal category beyond rote learning. This is all the more remarkable as the stimulus dataset used includes multiple animal species (mammals, reptiles, fishes, birds, etc.) and multiple factors that can affect the appearance of the target object category such as changes in position, scale, view-point or the type of background scenery.

A major challenge associated with the categorization of natural scenes is the presence of significant clutter. Previous psychophysics work has shown that the accuracy of human participants rapidly degrade in the presence of increasing clutter (Serre et al., 2007). We thus assessed next the robustness of the neural decoding to the presence of background clutter.

Robustness of the neural decoding to background clutter

The accuracy of neural decoding was further broken down according to different clutter conditions (Fig. 5). The accuracy of neural decoding appeared quite stable across clutter conditions and remained well above chance except for the most cluttered condition (where the animals tend to occupy a much smaller portion of space than background clutter).

Next, we linked early ventral stream neural activity to behavioral responses.

Fast single-trial decoding is predictive of behavioral responses

We computed an "animalness" accuracy score for individual images based on either monkeys' behavioral responses or neural decoding (Fig. S4). Such a score provides a compact characterization of the visual strategy employed by a visual system and permits the comparison between neural and behavioral data via direct correlation. The correlation between animalness scores computed from neural decoding (90–140 ms time window) and behavioral responses was significant (M1: $r^2 = 0.45$, $p < 0.001$; M2: $r^2 = 0.27$, $p < 0.001$), even after correcting for classification accuracy using partial correlation (see

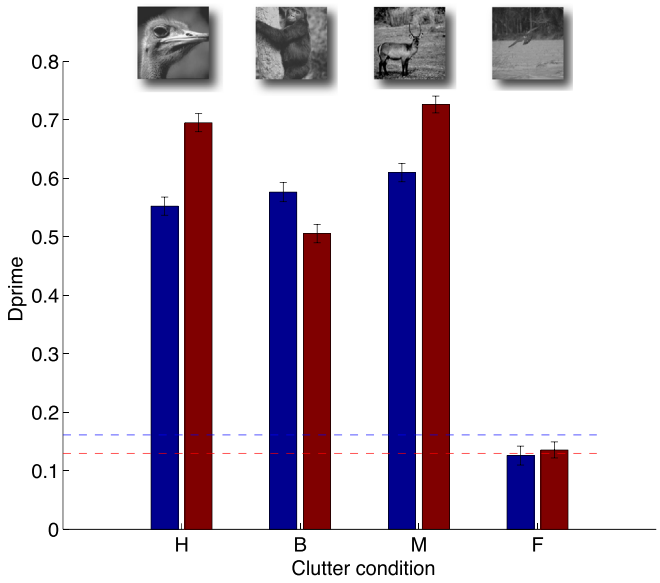

Fig. 5. Category information and clutter condition. Neural decoding (90–140 ms time window; 95% CI shown as error bars) for monkey M1 (blue) and M2 (red) shown for each sub-categories of clutter conditions (H: head; B: Body; M: Medium far and F: far). For more details on the stimuli see Serre et al. (2007). The horizontal dashed lines represent chance level for each monkey.

Methods; M1: $r^{2*} = 0.29$, $p < 0.001$; M2: $r^{2*} = 0.15$, $p < 0.001$). For comparison, we computed a similar score for both a representative feedforward computational model of the ventral stream of the visual cortex (HMAX) previously shown to match human performance on the same rapid animal categorization task (Serre et al., 2007) and for a low-level visual representation based on pixel intensities (see Supplementary Information).

To test for a more direct link between this early neural activity and behavioral response times, a hallmark of decision making processes (Johnson and Olshausen, 2003; Schall, 2002), we organized trials into quartiles based on the overall distribution of reaction times (Fig. 6a). Global decoding (90–140 ms post-stimulus onset time window) revealed a monotonic (near-linear) relationship between decoding accuracy and mean reaction times for each quartile: Trials corresponding to faster quartiles were decoded with higher accuracy (M1: $F(3,1196) = 1477$, $p < 0.001$; M2: $F(3,1196) = 820$, $p < 0.001$). Temporal decoding conducted on individual quartiles separately (Fig. 6b) further suggested that decoding latencies followed response times very closely. Trials associated with faster quartiles were decoded faster and with higher accuracy. These results demonstrate that fast neural activity in the ventral stream is linked to both behavioral accuracy and reaction time.

To further quantify the visual information encoded in this neural activity beyond task-related category information, we next assessed whether task-irrelevant visual information could still be decoded independently of monkey behavior.

Fine visual information can be decoded independently of monkey behavior

Here, we considered four (basic) categories of images from a subset of the target stimuli presented (people, macaques, chimpanzees, otter faces; Fig. 7). Reliable decoding of these subcategories above chance level suggests that task-related information does not completely override visual information. These results further suggest that, in principle, fast ventral stream neural activity could subserve multiple visual recognition tasks, consistent with earlier predictions from computational models (Joyce and Cottrell, 2004; Riesenhuber and Poggio, 1999).

Next, we evaluated the robustness of the neural decoding to the presence of background clutter.

Human and non-human primates share a similar visual strategy

Here, we compare the accuracy (d′) of the two monkeys to that of human participants (Serre et al., 2007) using a very similar paradigm with the same novel set of stimuli. Despite the presence of the mask, monkeys reached an accuracy level (M1: 1.35; M2: 1.92) comparable to that of human participants ($n = 22$; H: mean 1.96, SD = 0.50) for the same novel set of images.

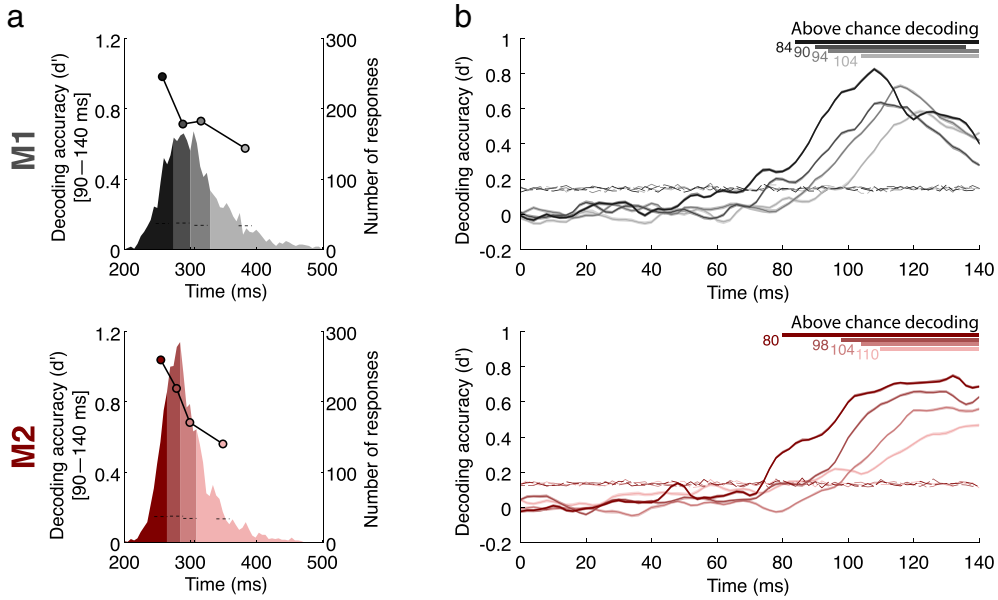

Fig. 6. Linking rapid ventral stream neural activity with response times. a) Trial binning according to reaction times and corresponding neural decoding (overlaid on the distributions) shown as circles (90–140 ms time window; 95% CI shown as error bars) for monkey M1 (gray) and M2 (red). Horizontal dotted lines indicate upper limit of the 95% CI around chance level obtained with a permutation procedure. b) Decoding conducted on individual quartiles. Center lines correspond to the average decoding accuracy estimated over multiple cross-validations of the data. Corresponding CIs (95%) obtained via bootstrapping are shown with transparency (most confidence intervals are actually too small to be visible). Significant deviation from chance level is shown at the top of the graph with horizontal bars along with the corresponding earliest latencies. Horizontal dotted curves indicate upper limit of the 95% CI around chance level obtained with a permutation procedure.

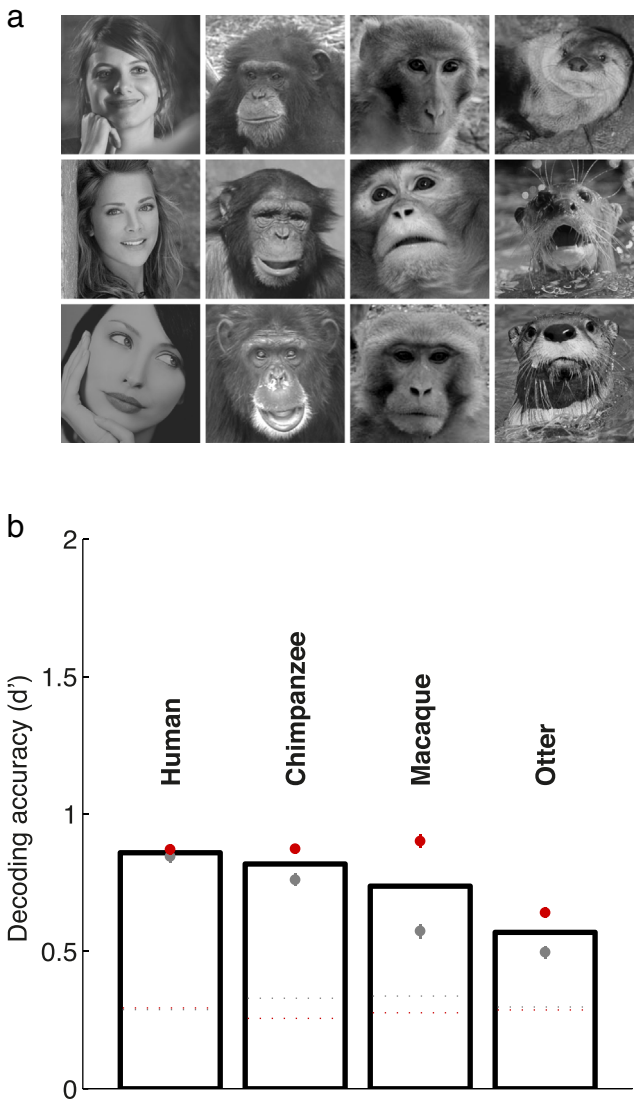

Fig. 7. Decoding of fine-level visual information. a) Sample stimuli used for the analysis. b) Accuracy for decoding animal sub-categories (human, chimpanzee, macaque and otter) for target images. Decoding accuracy (90–140 ms time window) based on a one-versus-all classification procedure for M1 (gray) and M2 (red). Bar plots indicate the average decoding for both monkeys. Horizontal dotted lines indicate the upper limit of the 95% CI around chance level obtained with a permutation procedure.

To further assess the similarity of the visual strategies used by monkeys and human participants, we computed animalness scores from behavioral responses for both monkeys and human participants (Fig. 8). For human observers, this score was computed as the fraction of human observers that classified a specific image as an animal irrespective of the actual category label. A score of 1.0/0.0 means that all participants classified the image as animal/non-animal. Any value in between reflects some variability across subjects. We computed a similar index for monkeys by pooling responses between the two animals and over multiple trials to obtain a reliable estimate.

We found a significant correlation between human observers and monkeys (Spearman correlation: $r^2 = 0.73$, $p < 0.001$) even after factoring out accuracy using a partial correlation measure ($r^{2*} = 0.33$, $p < 0.001$). This interspecies correlation was as strong as the correlation between the two monkeys ($r^2 = 0.67$, $p < 0.001$; partial correlation: $r^{2*} = 0.32$, $p < 0.001$). Overall this suggests that monkey and human participants do indeed follow a very similar visual strategy in our task.

Discussion

The present study investigated the neural underpinning of rapid natural scene categorization in non-human primates. Despite the inherent complexity of natural scenes, we found that superordinate object category can be read out very rapidly (within ~ 100 ms post stimuli onset) from intermediate areas (V4/PIT) of the ventral stream of the visual cortex.

One limitation of ECoG recordings is the lack of a precise spatial localization of the underlying neural source. We remain confident, however, that the recorded neural signals were not contaminated by motor preparatory responses because our analysis time window did not overlap with behavioral responses. Furthermore, in addition to task-relevant category signals, we were also able to decode task-irrelevant visual information and to map out RFs for individual electrodes.

Using backward masking, we found a striking degree of similarity between ventral stream neural activity and behavioral responses: both remained unaffected during an initial visual processing stage with the mask only impacting later processing. These results are consistent with earlier electrophysiological masking studies (Keysers et al., 2001; Kovacs et al., 1995; Rolls and Tovee, 1994) and masking theories (Breitmeyer, 2007; Breitmeyer and Ogmen, 2006) positing that visual processing of the stimulus and mask are kept separate during an initial short time period. These results are also consistent with theories postulating two distinct modes of visual processing, i.e., an early (possibly feedforward) processing unaltered by the presentation of a backward mask, which only interferes with later (feedback) processing (Lamme and Roelfsema, 2000; Schmidt and Schmidt, 2009; VanRullen and Koch, 2003). At the same time, given our relatively short estimate of the timing of the mask effect (~130 ms) compared to estimates of V4 attentional modulation reported in the literature (~160–170 ms, see (Buffalo et al., 2010; Poort et al., 2012)), the presentation of the mask in our study is likely to already interfere with feedforward processing. In addition, the observed mask timing does not exclude a possible fast recurrent modulation during the initial processing time window (Buffalo et al., 2010; Hupe et al., 2001). A more direct test for teasing apart alternative theories of visual masking would, however, require to vary the SOA more systematically to demonstrate co-variations between SOA, neural decoding and behavior, or the use of an inactivation protocol as done in Hupe et al. (1998).

Most previous attempts to link visual processes to behavior have focused on the processing of motion information in the dorsal stream of the visual cortex (Gold and Shadlen, 2007). A previous study based on single unit activity in IT did not find any co-variation between neural latencies in IT and reaction time (DiCarlo and Maunsell, 2005). The present study using ECoG electrodes, however, was able to identify co-variations between ventral stream neural activity and both perceptual decisions and their timing. This is consistent with a more recent single-cell study which demonstrated a correspondence between neural activity in IT and the speed of recognition (Mruczek and Sheinberg, 2007) using isolated objects in a visual search display. Furthermore, our estimate of the latency of category information agreed well with previous estimates on the optimal timing for IT micro-stimulations to affect perceptual decisions (Afraz et al., 2006).

Our analysis further suggests that relatively modest shifts in the latency of task-related information encoded in the ventral stream (4–18 ms) yields larger shifts in the corresponding distribution of behavioral responses (10–60 ms). One possible interpretation based on computational models of decision making such as the Drift Diffusion Model (DDM; (Ratcliff and McKoon, 2008)) is that ventral stream neural activity reflects the rate of accumulation of information, known as the drift rate, which is determined by the quality of the information extracted from the visual stimulus. The role of the ventral stream would thus be to convey the amount of evidence in the stimulus toward the target or the distractor decision category (Ratcliff and McKoon, 2008).

The impressive speed of visual processing observed during rapid categorization tasks has led some researchers to argue for subcortical

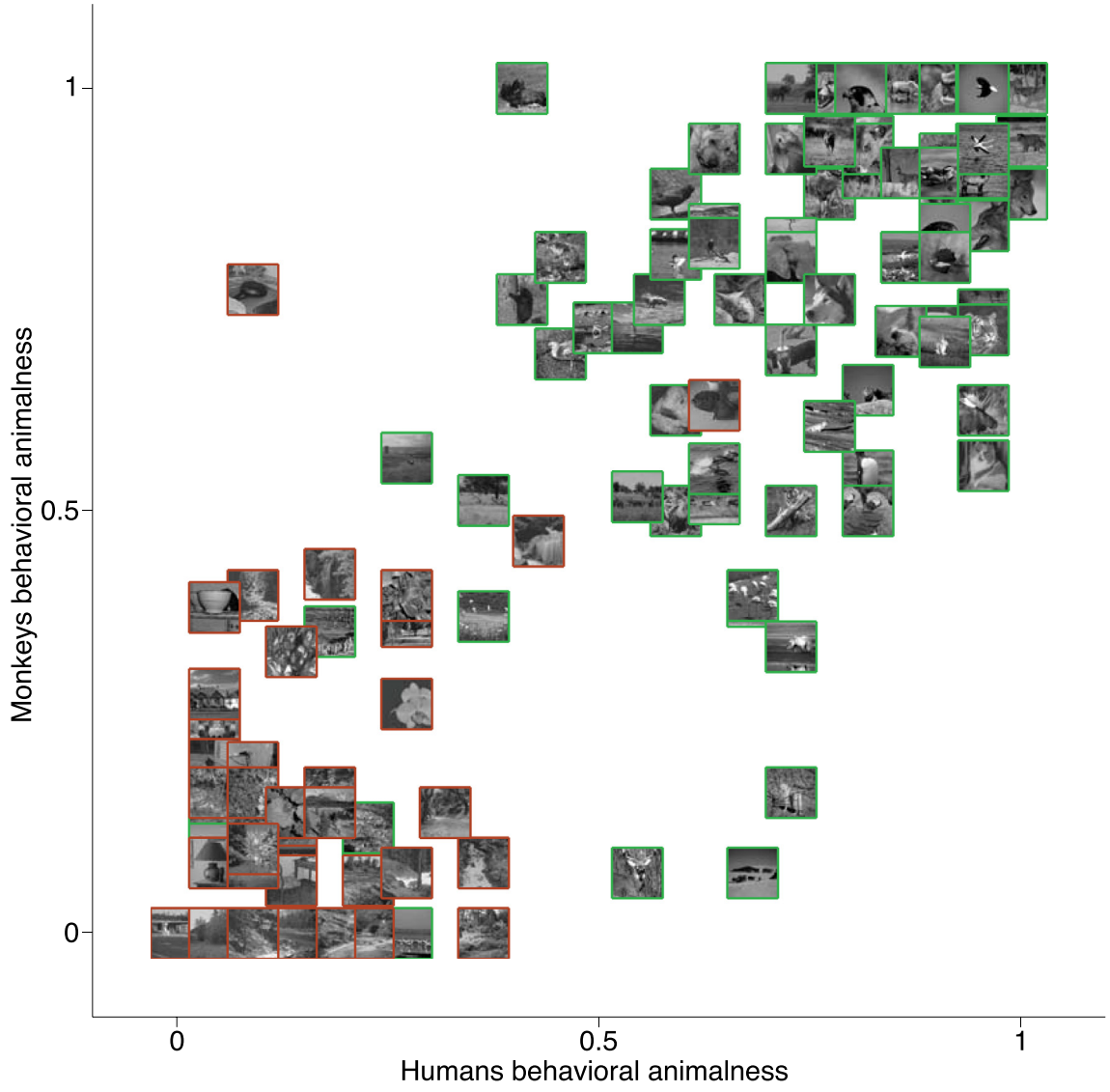

Fig. 8. Comparison between the visual strategy used by humans vs. monkeys. An animalness index was computed for individual images (120 animals and 120 non-animals) based on the fraction of trials that was classified as animal target (irrespective of the actual true label) by either human observers (x-axis; n = 22) or monkeys (y-axis; n = 2): the higher the score the more likely the image is to be classified as containing an animal. Thumbnails with green/red outlines correspond to animal/and non-animal stimuli.

routes bypassing altogether the visual cortex, e.g., via the thalamus through the amygdala (LeDoux, 1996). Direct projections between these two structures were observed on rodent electrophysiology during fear conditioning, but evidence for this "low-road" subcortical pathway for rapid vision is a matter of debate (see Cauchoix and Crouzet, 2013 for a recent review). The existence of animal category selective responses has been demonstrated in the human amygdala (Mormann et al., 2011). The observed latencies are relatively late (>300 ms) compared to those found in the present study (<100 ms). Interestingly a recent monkey electrophysiology study has shown selectivity for threatening stimuli (snakes) in the pulvinar within ~50 ms post-stimulus onset (Van Le et al., 2013). The present study supports more directly the "high-road" (cortical) hypothesis. At the same time, the existence of direct projections between the pulvinar and intermediate areas of the ventral stream of the visual cortex leaves, however, open the possibility that the visual selectivity for animal vs. non-animal found in the present study originates in subcortical areas with the ventral stream simply relaying the information to downstream areas (Pessoa and Adolphs, 2011).

Evidence suggesting that monkeys are capable of learning a high-level visual concept of abstract natural categories has so far been limited to behavioral studies (Fabre-Thorpe et al., 1998; Fize et al., 2011; Sigala, 2009). Previous monkey electrophysiology studies on categorization (e.g., cats vs. dogs (Freedman et al., 2001) or faces vs. non-faces and fishes vs. non-fishes (Sigala and Logothetis, 2002)) did not test for generalization to novel (unfamiliar) stimuli and are thus compatible with rote learning of individual exemplars (Vogels, 1999a, 1999b; Sigala et al., 2002). The present study establishes that ventral stream neural activity, in principle, could be supporting primates' generalization ability during high-level processing of natural scenes. In addition, we also found that it was possible to decode finer level (basic) category information, which was unrelated to the task (from target stimuli only). This suggests that, while salient, task-related category information does not completely override the visual information contained in these relatively early visual processes.

One of the main challenges associated with rapid natural scene categorization is clutter. Previous monkey electrophysiology studies (De Baene et al., 2007; Reynolds and Chelazzi, 2004; Rolls and Tovee, 1995; Sato, 1989; Sheinberg and Logothetis, 2001; Zhang et al., 2011; Zoccolan et al., 2005) have shown that ventral stream neural selectivity degrades when multiple objects are presented simultaneously. At the same time, the range of clutter tolerance found within and across

studies is quite broad (Zoccolan et al., 2007) and, ECoG decoding from human inferotemporal cortex is robust when two objects are presented simultaneously (Agam et al., 2010; Reddy and Kanwisher, 2007). Our study further suggests that, consistent with behavior, category decoding from ventral stream neural activity is at least partially robust to natural background clutter.

In addition, the present study found a high degree of correlation between neural and behavioral data as well as with a computational model of the ventral stream of the visual cortex. While we controlled for the most obvious low-level visual differences between the target and distractor set (e.g., distance to the camera, pixel intensities), we cannot rule out the possibility that relatively low-level features (including spatial frequencies) may be driving these correlations.

We further demonstrated that monkey and human participants exhibit similar patterns of correct and incorrect responses on the same set of images suggesting that they engage similar visual representations. These behavioral results adds to a growing body of evidence (Fize et al., 2011) suggesting that the neural mechanisms supporting rapid object categorization may be conserved between humans and macaque monkeys.

In sum, the present study suggests that rapid ventral stream neural activity induces a selective task-relevant signal subsequently used to drive visual categorization.

Acknowledgment

We would like to thank several of our colleagues for their valuable inputs on this manuscript: N. Bichot, D. Brooks, M. Franck, M. Peelen, T. Poggio, D. Sheinberg, I. Sofer and R. VanRullen. The monkey electrophysiology was supported by the FRM and DGA. The analysis of the data and comparison to models was supported by NSF early career award (IIS-1252951) and DARPA grant (N10AP20013) to T.S. and support through the Labex IAST (ANR-11-IDEX-0002-02) to M.C. Additional support to T.S. was provided by ONR grant (N000141110743), the Center for Computation and Visualization at Brown University, and the Robert J. and Nancy D. Carney Fund for Scientific Innovation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.10.012.

References

Afraz, S.-R., Kiani, R., Esteky, H., 2006. Microstimulation of inferotemporal cortex influences face categorization. Nature 442, 692–695.
Agam, Y., Liu, H., Papanastassiou, A., Buia, C., Golby, A.J., Madsen, J.R., Kreiman, G., 2010. Robust selectivity to two-object images in human visual cortex. Curr. Biol. 20, 872–879.
Bacon-Macé, N., Macé, M.J.-M., Fabre-Thorpe, M., Thorpe, S.J., 2005. The time course of visual processing: backward masking and natural scene categorisation. Vis. Res. 45, 1459–1469.
Biederman, I., 1972. Perceiving real-world scenes. Science 177, 77–80.
Breitmeyer, B.G., 2007. Visual masking: past accomplishments, present status, future developments. Adv. Cogn. Psychol. 3, 9–20.
Breitmeyer, B., Ogmen, H., 2006. Visual Masking: Time Slices through Conscious and Unconscious Vision. Oxford University Press.
Buffalo, E.A., Fries, P., Landman, R., Liang, H., Desimone, R., 2010. A backward progression of attentional effects in the ventral stream. Proc. Natl. Acad. Sci. U. S. A. 107, 361–365. http://dx.doi.org/10.1073/pnas.0907658106.
Cauchoix, M., Crouzet, S.M., 2013. How plausible is a subcortical account of rapid visual recognition? Front. Hum. Neurosci. 7.
Cauchoix, M., Arslan, A.B., Fize, D., Serre, T., 2012. The neural dynamics of visual processing in monkey extrastriate cortex: a comparison between univariate and multivariate techniques. Machine Learning and Interpretation in Neuroimaging. Springer, pp. 164–171.
Cauchoix, M., Barragan-Jason, G., Serre, T., Barbeau, E.J., 2014. The neural dynamics of face detection in the wild revealed by MVPA. J. Neurosci. 34, 846–854.
Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. TIST 2, 27.
Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. Nat. Neurosci. 17, 455–462.
Crouzet, S.M., Kirchner, H., Thorpe, S.J., 2010. Fast saccades toward faces: face detection in just 100 ms. J. Vis. 10.
De Baene, W., Premereur, E., Vogels, R., 2007. Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. J. Neurophysiol. 97, 2900–2916.
Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21.
DiCarlo, J.J., Maunsell, J.H., 2005. Using neuronal latency to determine sensory–motor processing pathways in reaction time tasks. J. Neurophysiol. 93, 2974–2986.
Eifuku, S., De Souza, W.C., Tamura, R., Nishijo, H., Ono, T., 2004. Neuronal correlates of face identification in the monkey anterior temporal cortical areas. J. Neurophysiol. 91, 358–371.
Evans, K.K., Treisman, A., 2005. Perception of objects in natural scenes: is it really attention free? J. Exp. Psychol. Hum. Percept. Perform. 31, 1476.
Fabre-Thorpe, M., 2003. Visual categorization: accessing abstraction in non-human primates. Philos. Trans. R. Soc. Lond. B Biol. Sci. 358, 1215–1223.
Fabre-Thorpe, M., 2011. The characteristics and limits of rapid visual categorization. Front. Psychol. 2.
Fabre-Thorpe, M., Richard, G., Thorpe, S.J., 1998. Rapid categorization of natural images by rhesus monkeys. Neuroreport 9, 303–308.
Fize, D., Cauchoix, M., Fabre-Thorpe, M., 2011. Humans and monkeys share visual representations. Proc. Natl. Acad. Sci. 108, 7635–7640.
Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2001. Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291, 312–316.
Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. Annu. Rev. Neurosci. 30, 535–574.
Hung, C.P., Kreiman, G., Poggio, T., DiCarlo, J.J., 2005. Fast readout of object identity from macaque inferior temporal cortex. Science 310, 863–866.
Hupe, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P., Bullier, J., 1998. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 394, 784–787.
Hupe, J.-M., James, A.C., Girard, P., Lomber, S.G., Payne, B.R., Bullier, J., 2001. Feedback connections act on the early part of the responses in monkey visual cortex. J. Neurophysiol. 85, 134–145.
Isik, L., Meyers, E.M., Leibo, J.Z., Poggio, T., 2014. The dynamics of invariant object recognition in the human visual system. J. Neurophysiol. 111, 91–102.
Johnson, J.S., Olshausen, B.A., 2003. Timecourse of neural signatures of object recognition. J. Vis. 3.
Joyce, C.A., Cottrell, G.W., 2004. Solving the visual expertise mystery. Prog. Neural Process. 15, 127–136.
Keysers, C., Xiao, D.-K., Földiák, P., Perrett, D.I., 2001. The speed of sight. J. Cogn. Neurosci. 13, 90–101.
Kirchner, H., Barbeau, E.J., Thorpe, S.J., Régis, J., Liégeois-Chauvel, C., 2009. Ultra-rapid sensory responses in the human frontal eye field region. J. Neurosci. 29, 7599–7606.
Kovacs, G., Vogels, R., Orban, G.A., 1995. Cortical correlate of pattern backward masking. Proc. Natl. Acad. Sci. 92, 5587–5591.
Kreiman, G., Hung, C.P., Kraskov, A., Quiroga, R.Q., Poggio, T., DiCarlo, J.J., 2006. Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. Neuron 49, 433–445.
Lamme, V.A., Roelfsema, P.R., 2000. The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci. 23, 571–579.
LeDoux, J., 1996. The Emotional Brain: The Mysterious Underpinnings of Emotional Life. Blockman, Inc., N.Y.
Liu, H., Agam, Y., Madsen, J.R., Kreiman, G., 2009. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 62, 281–290.
Macé, M.J.-M., Richard, G., Delorme, A., Fabre-Thorpe, M., 2005. Rapid categorization of natural scenes in monkeys: target predictability and processing speed. Neuroreport 16, 349–354.
Meyers, E.M., Kreiman, G., 2012. Tutorial on pattern classification in cell recording. Vis. Popul. Codes 517–538.
Mormann, F., Dubois, J., Kornblith, S., Milosavljevic, M., Cerf, M., Ison, M., Tsuchiya, N., Kraskov, A., Quiroga, R.Q., Adolphs, R., 2011. A category-specific response to animals in the right human amygdala. Nat. Neurosci. 14, 1247–1249.
Mruczek, R.E., Sheinberg, D.L., 2007. Activity of inferior temporal cortical neurons predicts recognition choice behavior and recognition time during visual search. J. Neurosci. 27, 2825–2836.
Pessoa, L., Adolphs, R., 2011. Emotion and the brain: multiple roads are better than one. Nat. Rev. Neurosci. 12, 425-425.
Poort, J., Raudies, F., Wannig, A., Lamme, V.A.F., Neumann, H., Roelfsema, P.R., 2012. The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. Neuron 75, 143–156. http://dx.doi.org/10.1016/j.neuron.2012.04.032.
Potter, M.C., 2012. Recognition and memory for briefly presented scenes. Front. Psychol. 3.
Potter, M.C., Levy, E.I., 1969. Recognition memory for a rapid sequence of pictures. J. Exp. Psychol. 81, 10.
Ratcliff, R., McKoon, G., 2008. The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput. 20, 873–922. http://dx.doi.org/10.1162/neco.2008.12-06-420.
Reddy, L., Kanwisher, N., 2007. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr. Biol. 17, 2067–2072.
Reynolds, J.H., Chelazzi, L., 2004. Attentional modulation of visual processing. Annu. Rev. Neurosci. 27, 611–647.
Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. Nat. Neurosci. 2, 1019–1025.

Rolls, E.T., Tovee, M.J., 1994. Processing speed in the cerebral cortex and the neurophysiology of visual masking. Proc. R. Soc. Lond. B Biol. Sci. 257, 9–16.
Rolls, E.T., Tovee, M.J., 1995. The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. Exp. Brain Res. 103, 409–420.
Rosch, E., 1975. Cognitive representations of semantic categories. J. Exp. Psychol. Gen. 104, 192.
Rousselet, G.A., Pernet, C.R., 2012. Improving standards in brain–behavior correlation analyses. Front. Hum. Neurosci. 6.
Sato, T., 1989. Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. Exp. Brain Res. 77, 23–30.
Schall, J.D., 2002. Decision making: neural correlates of response time. Curr. Biol. 12, R800–R801.
Schmidt, T., Schmidt, F., 2009. Processing of natural images is feedforward: a simple behavioral test. Atten. Percept. Psychophys. 71, 594–606.
Serre, T., Oliva, A., Poggio, T., 2007. A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. 104, 6424–6429.
Sheinberg, D.L., Logothetis, N.K., 2001. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. J. Neurosci. 21, 1340–1350.
Sigala, N., 2009. Natural images: a lingua franca for primates? Open Neurosci. J. 3, 48–51.
Sigala, N., Logothetis, N.K., 2002. Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 415, 318–320.
Sigala, N., Gabbiani, F., Logothetis, N.K., 2002. Visual categorization and object representation in monkeys and humans. J. Cog. Neurosci. 14 (2), 187–198. http://dx.doi.org/10.1162/089892902317236830.
Thompson, R.K., Oden, D.L., 2000. Categorical perception and conceptual judgments by nonhuman primates: the paleological monkey and the analogical ape. Cogn. Sci. 24, 363–396.
Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381, 520–522.
Van Le, Q., Isbell, L.A., Matsumoto, J., Nguyen, M., Hori, E., Maior, R.S., Tomaz, C., Tran, A.H., Ono, T., Nishijo, H., 2013. Pulvinar neurons reveal neurobiological evidence of past selection for rapid detection of snakes. Proc. Natl. Acad. Sci. 110, 19000–19005.
VanRullen, R., Koch, C., 2003. Visual selective behavior can be triggered by a feed-forward process. J. Cogn. Neurosci. 15, 209–217.
Vogels, R., 1999a. Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. Eur. J. Neurosci. 11, 1239–1255.
Vogels, R., 1999b. Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. Eur. J. Neurosci. 11, 1223–1238.
Yoshor, D., Bosking, W.H., Ghose, G.M., Maunsell, J.H., 2007. Receptive fields in human visual cortex mapped with surface electrodes. Cereb. Cortex 17, 2293–2302.
Zentall, T.R., Wasserman, E.A., Lazareva, O.F., Thompson, R.K., Rattermann, M.J., 2008. Concept learning in animals. Comp. Cogn. Behav. Rev. 3, 13–45.
Zhang, Y., Meyers, E.M., Bichot, N.P., Serre, T., Poggio, T.A., Desimone, R., 2011. Object decoding with attention in inferior temporal cortex. Proc. Natl. Acad. Sci. 108, 8850–8855.
Zoccolan, D., Cox, D.D., DiCarlo, J.J., 2005. Multiple object response normalization in monkey inferotemporal cortex. J. Neurosci. 25, 8150–8164.
Zoccolan, D., Kouh, M., Poggio, T., DiCarlo, J.J., 2007. Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. J. Neurosci. 27, 12292–12307.