HAL
open science

The characteristics and limits of rapid visual categorization.

Michèle Fabre-Thorpe

To cite this version:

Michèle Fabre-Thorpe. The characteristics and limits of rapid visual categorization.. Frontiers in Psychology, 2011, 2, pp.243. 10.3389/fpsyg.2011.00243 . hal-00633734

HAL Id: hal-00633734

https://hal.science/hal-00633734v1

Submitted on 19 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

frontiers in
PSYCHOLOGY

REVIEW ARTICLE
published: 03 October 2011
doi: 10.3389/fpsyg.2011.00243

The characteristics and limits of rapid visual categorization

Michèle Fabre-Thorpe*

CNRS, Université Paul Sabatier, Centre de Recherche Cerveau et Cognition, Toulouse, France

Edited by:
Gabriel Kreiman, Harvard, USA

Reviewed by:
Thomas J. Palmeri, Vanderbilt University, USA
Wolfgang Einhauser, Philipps-Universität Marburg, Germany
Jeff Bowers, University of Bristol, UK

Correspondence:
Michèle Fabre-Thorpe, CNRS CERCO UMR 5549, Pavillon Baudot – CHU Purpan, BP 25202, Toulouse Cedex, France.
e-mail: michele.fabre-thorpe@cerco.ups-tlse.fr

Visual categorization appears both effortless and virtually instantaneous. The study by Thorpe et al. (1996) was the first to estimate the processing time necessary to perform fast visual categorization of animals in briefly flashed (20 ms) natural photographs. They observed a large differential EEG activity between target and distracter correct trials that developed from 150 ms after stimulus onset, a value that was later shown to be even shorter in monkeys! With such strong processing time constraints, it was difficult to escape the conclusion that rapid visual categorization was relying on massively parallel, essentially feed-forward processing of visual information. Since 1996, we have conducted a large number of studies to determine the characteristics and limits of fast visual categorization. The present chapter will review some of the main results obtained. I will argue that rapid object categorizations in natural scenes can be done without focused attention and are most likely based on coarse and unconscious visual representations activated with the first available (magnocellular) visual information. Fast visual processing proved efficient for the categorization of large superordinate object or scene categories, but shows its limits when more detailed basic representations are required. The representations for basic objects (dogs, cars) or scenes (mountain or sea landscapes) need additional processing time to be activated. This finding is at odds with the widely accepted idea that such basic representations are at the entry level of the system. Interestingly, focused attention is still not required to perform these time consuming basic categorizations. Finally we will show that object and context processing can interact very early in an ascending wave of visual information processing. We will discuss how such data could result from our experience with a highly structured and predictable surrounding world that shaped neuronal visual selectivity.

Keywords: rapid categorization, natural scenes, early visual processing, objects, scenes

INTRODUCTION

How long does it take to process a natural scene? What kind of visual information can be extracted from a natural stimulus in the first few hundred milliseconds? How fast are the complex cognitive operations required for object categorization at different levels? Is there a hierarchical organization from detection to fine categorization? In the last two decades a large debate has engaged about such questions and is far from finding an end.

Visual categorization is a fascinating cognitive operation. The categorization of mammals, fish, birds, insects, or snakes into an "animal" category implies the grouping of objects despite very large physical differences. Moreover categorization can be made at different levels from the largest categories such as animal to the recognition of a "single item." An animal can be a dog, a greyhound, an afghan greyhound, or even Toby, my own afghan greyhound. Although such processing appears both effortless and virtually instantaneous in our daily life, the relative processing time necessary to complete such "cognitive" operations at different levels and the underlying cerebral mechanisms are still to determined.

How we move gaze to explore natural scenes has been subject of interest since the 1930s (Buswell, 1935). Yarbus's (1961) classic study on the ocular scanning of a scene showed that we typically make about three saccades a second, implying that a few hundred milliseconds is enough for visual analysis at fixation and for programming the next eye movement. A decade later, both Molly Potter (Potter and Levy, 1969; Potter, 1976) and Biederman (1972) showed that we could extract the gist of a scene from briefly glimpsed scenes, even at presentation rates of around 10 frames/s, although more processing time is needed to memorize pictures. But considering the massive parallelism in visual processing, various scenes could be processed simultaneously within different cortical areas and such studies do not really "time" the underlying processing.

Behavior is probably an unquestionable way to determine processing speed. If a reliable behavioral response to specific categories of stimuli can be performed with a particular floor latency, one can only conclude that the input–output loop can be completed in that time on at least some of the trials. The study by Thorpe et al. (1996) was the first to give a direct estimate of the processing time necessary to perform fast visual categorizations. Subjects were required to perform an "animal/non-animal" go/no-go categorization task as fast and as accurately as possible when presented with natural scenes that were flashed for only 20 ms to prevent any exploratory eye movements but left unmasked. The authors showed that the earliest behavioral responses could be produced in under 300 ms

and that a large differential EEG activity between correct target and distracter trials developed from 150 ms after stimulus onset. For the first time it was possible to put a value on the shortest processing time needed to perform a rapid visual categorization task. The strong temporal constraint set by the short 150 ms latency of the differential EEG activity led to the straightforward conclusion that when producing their fastest responses, subjects must be primarily relying on the first massively parallel and essentially feed-forward sweep of activity through the ventral visual pathway. Indeed, object categorization presumably relies on the processing of visual information involving all the visual areas (V1-V2-V4-PIT-AIT) along the ventral pathway (Thorpe and Fabre-Thorpe, 2001, 2003), and processing in each cortical area almost certainly involves more than one layer of synapses with time devoted to axonal conduction and synapse integration (Nowak and Bullier, 1997; Girard et al., 2001).

Since 1996, a large number of studies have been conducted by different groups, both in humans and monkeys, to determine the different characteristics of visual categorization and fast visual processing. I will review some of these studies and argue that fast visual processing has got limits because it is most likely based on coarse and unconscious visual representations automatically activated by the first available "magnocellular" information and that for such fast functioning, the visual system has been shaped by experience, extracting, and using the regularities of the surrounding visual world.

A VERY ROBUST "MINIMAL REACTION TIME"

In the Thorpe et al. (1996) paper, the notion of "minimal reaction time" (MinRT) in fast visual categorization was not yet made explicit. It is clear that very early behavioral responses can be anticipatory or can rely on "bets" made by the subjects. In fact the go/no-go rapid categorization task is a good task to overcome such problems. Since targets and distracters were equiprobable, all anticipatory responses should be equally distributed on target and distracter trials. As a result the MinRT can be defined as the first time bin for which correct responses start to significantly outnumber incorrect responses (Fabre-Thorpe et al., 1998; VanRullen and Thorpe, 2001a). At such latencies, the processing must be completed, at least in some cases, and this value can be considered as a floor limit for an input–output loop in the task at hand.

The fact is that this MinRT appears very robust to different types of object categories. Indeed animals can be considered as a very special object category: they are very biologically pertinent and evolution might have developed hardwired systems to deal with stimuli that could be life threatening as for predators (New et al., 2007). It was thus crucial to demonstrate whether this optimal processing could be seen for object categories that appeared more recently in our evolutionary history and for which only life-time experience could be involved. This demonstration was done by comparing fast visual categorization for animals and vehicles (VanRullen and Thorpe, 2001a) and showed that reaction time distributions were statistically indistinguishable in particular when focusing on early responses. Furthermore, MinRT could be evaluated at 250 ms arguing strongly that at least on some trials, the entire visuo-motor sequence (visual processing, decision process, and motor output) could be completed within this short temporal window. In that study, the MinRT was particularly short, but 250–290 ms appears as a value widely found for human subjects categorizing superordinate object classes. This is indeed the case when participants are required to categorize visual scenes as being natural or artificial (Joubert et al., 2007), or to categorize human being from animals (Rousselet et al., 2003). MinRT of 260 ms were also reported when subjects were required to randomly categorize a single image or two simultaneously flashed images (Rousselet et al., 2002).

Interestingly this MinRT is no shorter for classes of stimuli for which the literature might have predicted an advantage such as simple forms: squares vs. circle (Aubertin et al., 1999) or even for human faces among animal faces (Rousselet et al., 2003).

It might thus be that this value corresponds to a floor effect observed in each study because of some stimuli in the set that are so trivial to process that optimal processing speed is always reached (a single animal, in canonical view, well segregated on a uniform background for example). Optimal RTs could also have been expected in response to the presentation of a threatening animal (Ohman et al., 2001; Lobue and DeLoache, 2008). The first possibility was ruled out by tracking the effect of familiarity on fast categorization tasks (Fabre-Thorpe et al., 2001). Extensive training with a given set of natural scenes did not result in any processing speed increase for behavioral responses triggered before 360 ms (**Figure 1**). A training effect was observed for a set of "difficult stimuli" that could be evidenced because they induced more errors and long RTs when a correct go-response was produced. On the other hand whereas such a set of "difficult stimuli" could be clearly extracted, no set of "easy stimuli" could be demonstrated. But the distribution of images categorized with short RTs fitted exactly the prediction of a random distribution over the RT range, so that there was no evidence that some "trivial" stimuli were associated with particularly fast responses. The only evidence that we ever found was in favor of threatening animals (Delorme et al., 2010). Reptiles were categorized with faster median RTs but this effect was only global and could not be observed for behavioral responses produced at latencies earlier than about 340 ms.

The robustness of the MinRTs in most of the above studies was corroborated by the associated EEG recordings. Indeed the large differential EEG activity between target and distracter correct trials recorded over frontal and occipital sites developed at the same latency for animals or vehicles (VanRullen and Thorpe, 2001b), for familiar vs. novel stimuli (Fabre-Thorpe et al., 2001) and for one vs. two processed images (Rousselet et al., 2002).

Obviously, trying to optimize the top down presetting of the visual system by increasing the subjects familiarity with a set of pictures was not enough to increase processing speed. To investigate further how top down knowledge could influence the visual analysis of natural scenes, another go/no-go task required humans to categorize a single specific learned natural scene among a variety of different non-target other natural photographs seen for the first time. In this study (Delorme et al., 2004), maximal top down control could be used to "preset" all neurons selective for features diagnostic of the single target–scene. This was indeed the first time (**Figure 1**) that we found a decrease of MinRT (220 ms in the single target–scene task vs. 260 ms in a animal/non-animal

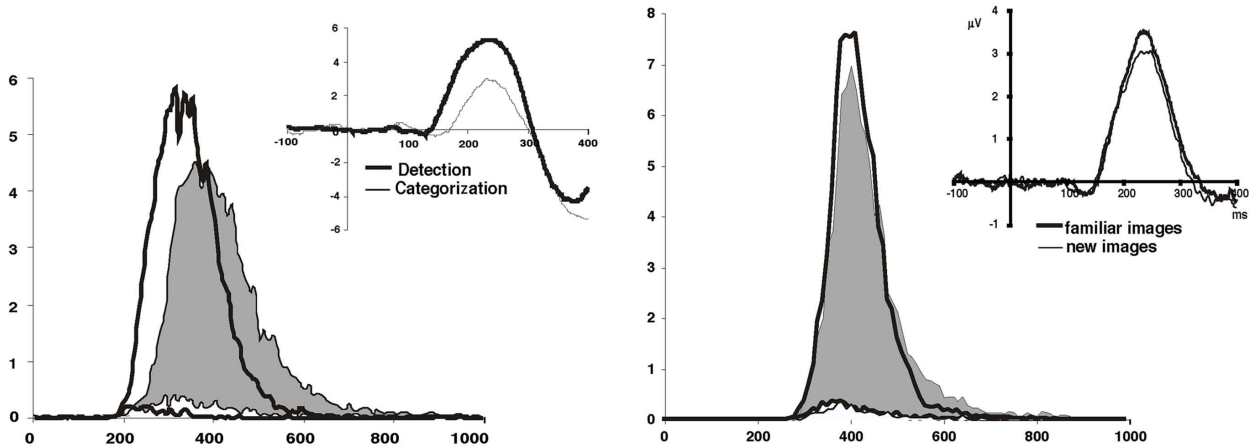

FIGURE 1 | Overall reaction time distribution of go-responses expressed in milliseconds and percentage of total go-responses using 10 ms time bins. In shaded gray and thin lines the results from the fast animal go/no-go categorization task. Top traces correct go-responses, bottom traces incorrect go-responses (false alarms). Comparison is made with the detection of a single target–scene (on the left) and with the same task using familiar images: images on which subjects had been daily trained for 3 weeks (on the right). Differential EEG activity between correct go and no go trials is shown for each of the two conditions in the corresponding insert using the grand average (14 subjects) obtained on electrode FZ. Data from Delorme et al. (2004) and from Fabre-Thorpe et al. (2001).

categorization task). Note also that this study shows that detecting a specific scene is faster than categorizing a scene.

If familiarity does not affect the MinRT by inducing an increase in processing speed, one might expect that longer RT latencies for early responses could be observed with unusual presentations such as inverted or rotated stimuli. There again, the robustness of this MinRT is astonishing. MinRT was only found to be delayed (by 10–20 ms) for inverted vs. upright human faces. No impact on fast processing speed could be demonstrated in the processing of inverted humans, animal faces, and animals even when they were presented in all possible orientations (Guyonneau et al., 2006).

Finally, the MinRT value is dependent of the species studied. Humans are not the only species able to perform a fast animal visual categorization task of natural photographs. Other non-human primates such as baboons, macaques, and even marmosets can categorize natural scenes depending on the presence of an exemplar object belonging to a given target category such as persons, animals, birds, trees. . . (Roberts and Mazmanian, 1988; Fabre-Thorpe et al., 1998; Vogels, 1999; Martin-Malivel and Fagot, 2001; Minamimoto et al., 2010). Using the same fast go/no-go animal categorization task with briefly flashed unmasked pictures that was used in humans, monkeys' accuracy scores are slightly lower than humans (90 vs. 94%) but the speed of response is much faster! Median RTs observed in monkeys can be 100–200 ms faster than for humans and MinRT for monkeys are observed at 180 ms (**Figure 2**). As a lot is known about the functional anatomy of the monkey visual system and the neuronal response latencies in cortical areas along the ventral visual pathway, the temporal constraints set by this MinRT emphasize the role played by the first feedforward wave of processing (Thorpe and Fabre-Thorpe, 2001) in fast visual tasks. Conduction speed being fairly slow between cortical areas and even slower within different layers of a given cortical area, the difference in response speed could simply be explained by the relative brain sizes of humans and monkeys.

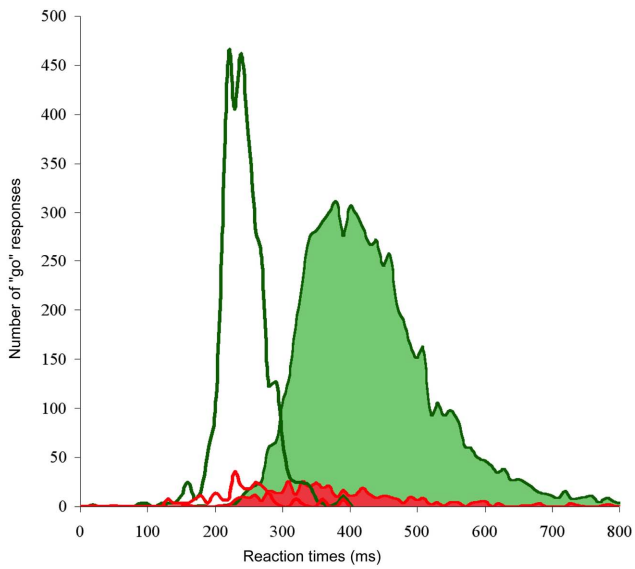

FIGURE 2 | Overall reaction time distribution of go-responses expressed in milliseconds and number of go-responses using 10 ms time bins. Human performance (shaded distributions) and monkey performance (empty traces) are shown in an animal/non-animal categorization task. Top traces: correct go-responses, bottom traces: incorrect go-responses (false alarms). Data from Fabre-Thorpe et al. (1998) and Delorme et al. (2000).

All the results obtained in the above mentioned studies emphasize that a minimum processing time cannot be shortened in fast visual categorization tasks even when processing stimuli – such as human faces – that are claimed to have a very special status because they could depend upon a specific module of processing or when processing the gist of a scene that could rely on global statistics of the photographs (Oliva, 2005; Oliva and Torralba, 2006).

To fit with such temporal constraints and reach optimal processing speed the visual system might rely on the first available visual information. This issue is considered in the next paragraph.

FAST VISUAL PROCESSING: A ROLE FOR THE MAGNOCELLULAR PATHWAY?

Visual information captured by retinal cells can reach cortical area V1 using either the magnocellular (M) system or the parvocellular (P) system. Traditionally, the M system has been associated with the dorsal visual stream and the extraction of structure from motion whereas the P system is associated with the ventral visual stream and with the fine analysis of static images. However this is an oversimplification as the M stream is also present in the ventral pathway and might even account for as much as half of the visual ventral information (Ferrera et al., 1992; Nealey and Maunsell, 1994). Visual information in the P stream reaches the visual cortex roughly 20 ms after M based information (Nowak et al., 1995; Nowak and Bullier, 1997) a temporal delay which means that if the visual system uses the first available information it has to rely on the magnocellular stream.

To disentangle the relative role of the two streams in fast visual categorization we have to rely on characteristics that are specific to either the M or the P pathway. For example, the P stream transmits chromatic information whereas the M information is motion and luminance based. In addition, the different visual pathways have different contrast sensitivities. Retinal and lateral geniculate parvocellular cells stop responding for contrast thresholds much higher than magnocellular cells (Kaplan and Shapley, 1986). Finally, magnocellular ganglion cells in the macaque retina are eight times less densely packed than parvocellular cells (Silveira and Perry, 1991), with more convergence from photoreceptors (Dacey and Petersen, 1992), so that magnocellular spatial resolution is relatively poor.

Together, all these differences between the M and P pathways allow predictions to be made and tested. If based on first available visual information, ultra-rapid categorization should rely on magnocellular object representations that will be very coarse, color-blind, and robust to contrast reductions. Indeed in everyday life, there are many conditions where visual conditions are often far from being optimal. Luminance and contrast can be very low, at dusk and dawn for example, and conditions might not always allow the processing of colors. However, even when facing such challenging everyday conditions our visual system still appears to perform very efficiently and without effort. The series of studies that were conducted both in humans and in monkeys were able to provide convincing arguments for all the above predictions.

FAST VISUAL CATEGORIZATION IS BASED ON ACHROMATIC OBJECT REPRESENTATIONS

Color has been shown to enhance recognition memory by conferring an advantage during encoding an retrieval in recognition memory paradigms such as delayed match to sample tasks of natural scenes (Gegenfurtner and Rieger, 2000; Wichmann et al., 2002; Spence et al., 2006). On the other hand we found that color had virtually no role in fast visual categorization tasks (Delorme et al., 2000, 2010). Indeed when monkeys or humans were performing a go/no-go animal/non-animal categorization task with a set of new natural scenes that could appear randomly in color or in black and white, no difference could be seen in response accuracy or speed of early responses. An influence of color is observed later, for responses with latencies over 325–400 ms.

Indeed, color could be used in later stage of processing (Yao and Einhauser, 2008; Elder and Velisavljevic, 2009). Colors could be used when diagnostic for the task at hand (Oliva and Schyns, 2000). They can even be used in very early processing, for example, in the task using a single target–scene described earlier (Delorme et al., 2004), spots of color were used by subjects when color was diagnostic of the target–scene.

The study by Delorme et al. (2000) showed that the early responses in fast visual object categorization task rely on achromatic representation of objects and thus provides a first argument in favor of a predominant role for magnocellular information to explain the fastest behavioral responses. The goal of the second series of studies discussed next and conducted in both humans and monkeys was to evaluate the robustness of fast visual categorization in challenging situations for the visual system when the contrast of achromatic images was dramatically reduced.

FAST VISUAL CATEGORIZATION IS VERY ROBUST AT VERY LOW CONTRAST

Here, the rational was to use the different contrast sensitivities of the M and P pathways. At very low residual contrasts, the parvocellular information is no longer available and the level of performance should only rely on magnocellular information. Both humans and monkeys were tested with achromatic images in which the original contrast of each scene was reduced by a factor of 2, 4, 8, 10, 16, or 32. This contrast reduction was done with mean luminance of the image kept constant and corresponded to a division of the SD of the pixel luminance values. If fast categorization relies only on parvocellular information, a drop of performance must suddenly appear when local contrast is below the parvocellular threshold (a value reported at 10% when calculated with black and white gratings). But performance was surprisingly robust in monkeys (Macé et al., 2010) and humans (Macé et al., 2005) that both showed a very similar progressive performance drop with decreasing contrasts (**Figure 3**). When 10% of the original contrast of the photographs was reached, thus well under the threshold of the parvocellular cells, performance was still at 70% correct and chance level was only observed when the original contrast was further divided by a factor of 32 (i.e., when photographs only contained 3% of the original contrast). At very low contrasts, performance has to rely only on the contrast sensitive magnocellular pathway and the robust performance evidenced both in humans and monkeys argue again for a strong contribution of the fast achromatic magnocellular information in rapid categorization tasks.

FAST VISUAL CATEGORIZATION AT THE CORNER OF THE EYE

Compared to parvocellular spatial resolution, magnocellular spatial resolution is relatively poor. Extrafoveal vision has a low spatial resolution and as a consequence, it is usually assumed that peripheral vision is inappropriate for object recognition. Indeed in everyday life when an object appears in far periphery we usually make a succession of eye movements to bring it in foveal vision. But what kind of decision can we take when forced to rely

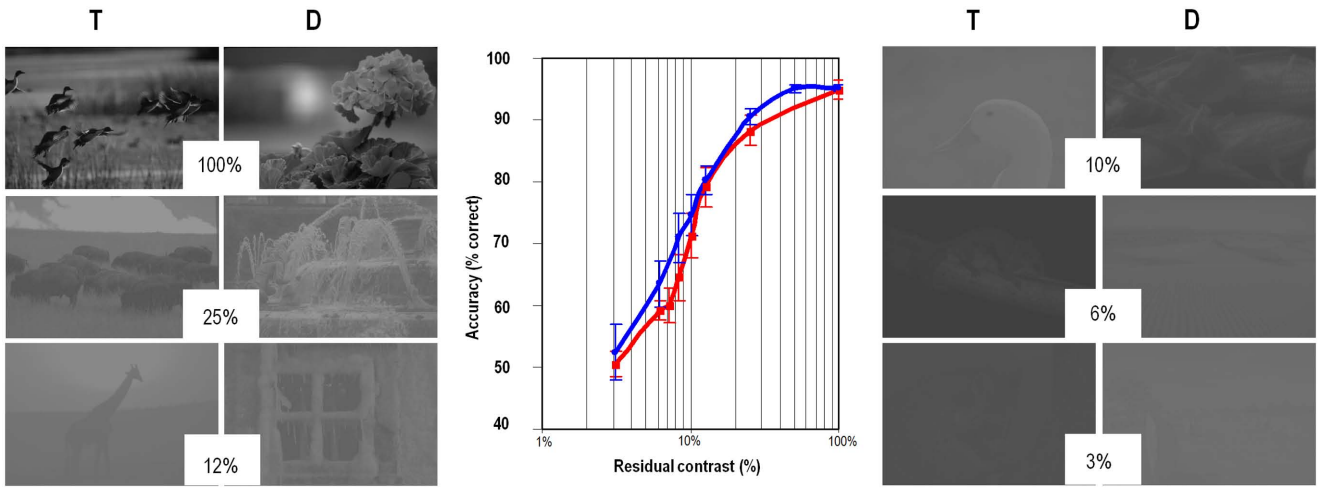

FIGURE 3 | Categorization accuracy (T for animal target and D for non-animal distracters) is robust to reduction of contrast. Around 94% correct with the original black and white image (considered as 100% contrast) accuracy was still around 70% with only 10% residual contrast for both humans (in red) and monkeys (in blue). Chance level (50%) was reached for both human and monkeys when only 3% of the initial contrast was preserved (adapted from Macé et al., 2010).

on peripheral vision? Whereas most studies investigated differences between central and peripheral vision in the processing of simple physical features such as contrast, size, spatial frequencies, phase, and texture discrimination, very few aimed at determining the abilities of peripheral vision in object recognition. There again, human performance levels are surprising (Thorpe et al., 2001). Subjects were required to perform an animal/non-animal fast visual categorization task. They looked straight ahead a fixation cross and large natural scenes were randomly flashed in nine positions virtually covering the whole horizontal visual field. The most peripheral scenes were centered at 57° and 70° and no effort was done to adjust the scaling of the image. Subjects were able to score respectively over 70 and 60% correct at these far eccentricities. The results demonstrate that high-level visual tasks such as object categorization can be performed from low spatial resolution information provided by the peripheral retina.

In that study, as for stimuli flashed with strong contrast reductions, the subjects were reporting that they were only guessing and would certainly not be able to report the animal that was presented suggesting that object representations were not accessible to consciousness (Boucart et al., 2010).

AN UNDERESTIMATED ROLE FOR THE MAGNOCELLULAR STREAM IN OBJECT PROCESSING

Back in the 80s, Sherman (1985) proposed that the parvocellular stream could provide high acuity descriptions to a coarse magnocellularly driven form of vision. But there are only a few studies that have explicitly considered such hypotheses (Kruger et al., 1988; Strasburger and Rentschler, 1996; Bullier, 2001; Macé et al., 2005, 2010). Neuronal responses provide arguments supporting early coarse information followed by more detailed object description. Indeed, Sugase et al. (1999) showed a biphasic response of IT neurons to faces with a first phasic component related to face recognition and a second late tonic component related to finer computations about facial characteristics (such as its expression). Some authors have proposed an influence of magnocellular information through projections of the dorsal visual cortical pathway over the ventral pathway (Vidyasagar, 1999; Bullier, 2001). However, as noted above, magnocellular projections might account for as much as half of the information in the ventral pathway. What we would propose is a rapid preprocessing of magnocellular inputs within the ventral pathway that would be able to guide, in an intelligent way, and through rapid short feed-back loops, the detailed visual processing of the slower parvocellular information (**Figure 4**).

The object representations built from early coarse magnocellular information might be sufficient to allow some forms of object categorization such as to detect an animal, a vehicle, a person, or a face in an image but may not be sufficient for finer discriminations. In that case, the processing time for a more restricted basic category of animals such as dogs or birds may need more information uptake and longer processing times; a prediction that is at odds with the widely accepted idea that basic representations are accessed first.

FROM SUPERORDINATE TO BASIC CATEGORIES: A LIMIT TO FAST VISUAL PROCESSING

As mentioned in the introduction, an object can be categorized at different levels: superordinate (animal, vehicle), basic (a dog, a car), subordinate level (a greyhound, a sports car), and even as a "unique item" Toby, my own afghan greyhound, or my Ferrari. The very predominant view is that the fastest level to be accessed is the basic level (Rosch et al., 1976). This view was further refined by Jolicoeur et al. (1984) who reported that some atypical members of basic level categories could be categorized faster at the subordinate level (ostrich vs. bird). More processing time would be needed for accessing large abstract superordinate categories or more detailed subordinate categories (Kosslyn et al., 1995; Jolicoeur et al., 1984;

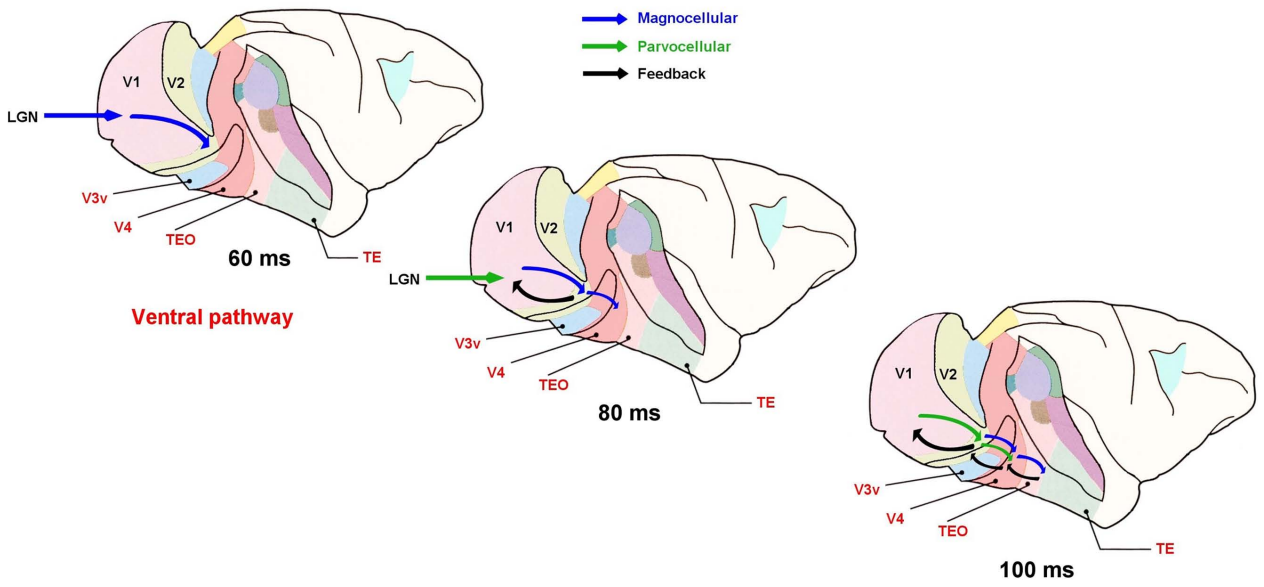

FIGURE 4 | A first wave of coarse achromatic magnocellular information (in blue) would first reach high-level cortical areas of the visual ventral pathways. At each step fast short feed-back loops (in black) could be used for guiding – in an intelligent way – the parvocellular information arriving with a temporal lag of about 20 ms. From Macé (2006) with permission.

Palmeri and Gauthier, 2004 for a review). But expertise can also play a role as experts can access both subordinate and basic category level at similar latencies (Tanaka and Taylor, 1991; Tanaka and Curran, 2001).

Given the temporal constraints set with fast visual categorization tasks using large superordinate categories, it seems difficult to predict even shorter MinRT for basic categories. To investigate the processing time necessary at the basic level of categorization, subjects were asked to perform rapid visual go/no-go categorization tasks using three different types of targets: animal (superordinate) and dogs or birds (basic levels). Natural images can induce biases so that we made sure that the performance in accuracy and speed on birds and dogs pictures when categorized at the superordinate "animal" level was similar to the values seen with other animals. The results of the study were really clear (Macé et al., 2009). To produce a correct go-response to a bird- or a dog-target an additional 40–65 ms of mean processing time was required (**Figure 5**). The whole reaction time distribution of correct go-responses was shifted toward longer latencies so that MinRT was also increased by 40–50 ms.

There is a large debate on the time needed to reach basic level object representation. Grill-Spector and Kanwisher (2005) claimed that "As soon as you know it is there, you know what it is," but subsequent studies claimed the contrary "Sometimes you know it is there before you know what it is" (Mack et al., 2008; Mack and Palmeri, 2010a), and "Detecting objects is easier than categorizing them" (Bowers and Jones, 2008). But when addressing categorization at the basic level, the nature of the distracter images is crucial in determining whether basic level categorization is truly required. As pointed out by Mack and Palmeri (2010a), the target vs. non-target tasks used by Grill-Spector and Kanwisher (2005) could sometimes be solved at the superordinate level

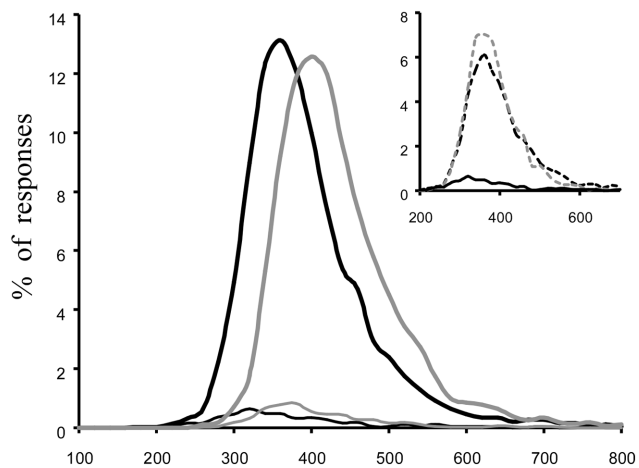

FIGURE 5 | Overall reaction time distribution of go-responses expressed in milliseconds and percentage of total go-responses using 10 ms time bins. Top traces correct go-responses, bottom traces incorrect go-responses. Comparison is made between the fast superordinate animal go/no-go categorization task (in black) and a fast basic "bird" categorization task (in gray). The insert compares the result obtained in the fast superordinate animal go/no-go categorization task when considering separately birds photographs (gray dotted line) and other non-birds animals (in black) showing that the two sets of images have similar difficulties when categorized at the superordinate level. Adapted from Macé et al. (2009).

(faces/cars for example). Even when the non-target images came from the same superordinate category they were restricted to very dissimilar non-target exemplars. Thus when dogs were targets the non-target images from the superordinate category were restricted to fish and birds – no other quadrupeds were used. In contrast,

Bowers and Jones (2008) used cats and dogs that are visually very much more similar In both cases dogs have to be categorized at the basic level because the non-target images are exemplars of the same superordinate category (animals), but the difficulty of the task is manipulated by mixing the dog targets among very dissimilar (birds, fish) or very similar (cats) non-target animals. In our 2009 study, we made sure that the non-target animals were very varied and found that categorization errors were biased toward more similar animals. With bird targets, errors were often made toward fish. In contrast, when dog were targets, errors were biased toward quadrupeds and mainly toward wolves, foxes, and bears. The result of basic level categorization of birds and dogs among a very varied selection of distracters in the animal kingdom was a clear shift in response latencies toward longer RTs.

When comparing superordinate and basic level in our study (Macé et al., 2009), animal-targets were presented among non-animal scenes, dog- and bird-targets among 50% non-animal scenes, and 50% other non-target animals of all kinds. But, what would happen to the processing time needed to reach basic level representation if the proportion of distracters containing a non-target animal dropped to 0%? In such case, we are back to a superordinate contrast (Bowers and Jones, 2008; Mack and Palmeri, 2010). Furthermore, the "target space" is restricted to "dogs" with the result that top down presetting can be maximal. In such conditions, we could expect the earliest responses to be triggered with shorter MinRTs and possibly find an advantage for the basic category. But when contrasting dogs with non-animal scenes, the results failed again to show any speed advantage at the basic level, emphasizing again the robustness of the MinRT value for object categorization. Basic dog categorization was performed with higher accuracy when compared to superordinate animal categorization (significant difference of about 4% correct) but median or MinRT values were strikingly similar. Even in such conditions, a basic level category is not accessed faster than its superordinate but at a similar latency (Macé et al., 2009).

The additional processing time needed to access basic level categories was also reproduced using scene gist categorization. Natural or man made superordinate scene categories were accessed much faster than basic level scene categories such as sea, mountains, street, or indoor scenes (Joubert et al., 2007). It is worth noticing that scene categorization (Natural/Man made) was not performed any faster than object categorization at the superordinate level (animal, vehicle. . .). Indeed various studies have suggested that a coarse lay-out of the scene could be available before object recognition takes place (Schyns and Oliva, 1994; Oliva and Torralba, 2001). The results obtained in our study argue more for a parallel processing of scene and object.

Some other recent studies also support the view of a faster access to superordinate coarse categories. The richness of our perception when we glance at a grayscale photograph of a natural scene was tested when the photograph was masked after a variable (27–500 ms) delay (Fei-Fei et al., 2007). Subjects were asked to report with accuracy what they had seen. "Animals" were reliably reported with shorter image presentations than "birds," "dogs," or "cats"; "vehicles" were reported before "cars"; "natural" and "man made" scenes were reported before "mountains," "forest," or "urban" scenes. The same temporal succession is found at the cerebral level, Martinovic et al. (2008) found that early (around 100 ms) synchronizations in brain activity might be sufficient for superordinate representation of objects in scenes, whereas entry level categorization would depend upon later (200–400 ms) brain activity.

These results are at odds with the predominant view that states basic category levels as special and accessed before superordinate or subordinate representations. How can we explain the clear advantage found for superordinate categories in fast categorization tasks?

One explanation might lie in the kind of task used in the different studies. The primacy of basic level categories has been reported in numerous studies involving naming tasks and verification tasks. In a verification task participants are first shown a category label ("animal," "dog," "greyhound") they are then shown the picture of an object, and they have to verify whether the object belong to the category label. Lexical access is thus necessary and it is well known that basic words are much more frequently used than superordinate words. A large temporal cost in terms of word retrieval or reference to lexical knowledge might therefore have masked the superordinate advantage when object processing is largely visual. Even when tasks have been carefully designed to try and differentiate lexical access from visual categorization (Tanaka and Taylor, 1991; Tanaka and Curran, 2001), there is no evidence that participants do not need to use lexical information during the course of the task, especially when target categories are changed from trial to trial. This is reflected in the much longer RTs usually reported for verification and naming tasks when compared to fast visual categorization in which subjects are required to take fast decisions. Such temporal constraints may partly explain our results (Rogers and Patterson, 2007). Moreover, the block procedure used in fast visual categorization allows the maximal "presetting" of the visual system for the task at hand. In such block procedures, motor responses could also rely on unconscious object representations, as suggested by some of our results (Thorpe et al., 2001), whereas activation of conscious representations might be needed in naming and verification tasks.

Another striking difference is the presentation time of the stimuli since the 20-ms flashed scenes used in ultra-rapid categorization tasks might favor coarse-level visual information because more fine-grained visual information would be degraded. Two points have to be emphasized. The first one concerns the fact that even if flashed, the stimuli are unmasked so that retinal persistence may be involved. Furthermore, even if information uptake is limited to 20 ms, information processing would be less disrupted than when strong masking is used. The duration of stimulus presentation might play a role, but recent – still unpublished – data suggest that the superordinate advantage persists even with longer stimulus presentation times.

Finally another difference concerns the stimuli used since most studies use drawings or objects shown in isolation. It could be the basic advantage is significantly reduced when the objects to be categorized are embedded in natural scenes (Murphy and Wisniewski, 1989).

For Jolicoeur et al. (1984), particular object category levels must be mapped into a single (or small set) of representations in memory. Such mapping would depend upon perceptual characteristics

(such as the object's shape) and cognitive factors (such as context) and would be faster at the entry level. But because the visual system is shaped by experience, neuronal selectivities for frequently encountered objects can be built progressively (Ahissar and Hochstein, 2004; Masquelier et al., 2008). What I would like to argue with others is that when such visually selective neurons exist, object processing is automatic, and can reach implicit object representation and bias behavior. Because of progressive integration of perceptual information over time, categorization levels are reached from coarse to fine, thus from superordinate to basic, and subordinate categories.

HOW AUTOMATIC IS ACCESS TO BASIC CATEGORIES?

The responses produced in the fast go/no-go categorization task could rely on early non-conscious object representations. Indeed, numerous studies have argued that implicit processing of objects can bias behavior and speed response production as in patients with prosopagnosia or unilateral neglect for example (De Haan et al., 1987; Driver and Vuilleumier, 2001; Forti and Humphreys, 2007). Object perception can potentiate appropriate actions toward that object (Grezes et al., 2003). One characteristic of the fast visual categorization task at superordinate level is that it can be performed in the near absence of attention. In 2002, two different studies reached this conclusion. The first one (Rousselet et al., 2002) showed that animal rapid categorization could be performed in parallel when subjects had to process two natural scenes flashed simultaneously, the second study used a dual task protocol, and showed that the animal and vehicle categorization could be performed in periphery without performance cost when attentional resources were fully engaged on the performance of a central attentionally demanding task (Li et al., 2002). Other studies using dual tasks further extended this finding and showed that gender categorization or even person identification could be done without focused attention (Reddy et al., 2004, 2006). Although the parallel processing of natural scenes has a limit (Rousselet et al., 2004; VanRullen et al., 2004), dual tasks can reveal which object category can be processed preattentively in such scenes. In fact VanRullen et al. (2004) proposed that the processing of "a feature, stimulus, or object category is preattentive if there exists a neuronal population selective to this feature, stimulus, or object category, independent of the cortical area involved."

Back to basic level object representations, the shift of response latencies toward longer reaction time values might be explained if the activation of these basic representations required attention. However, using a dual task protocol, we have recently demonstrated that even categorization at the basic level could be done without focused attention for both biological (dogs) and for man made (cars) objects (Poncet et al., 2011). Automatic processing of objects can thus reach a fine and detailed level of representation. Such data could fit with the reverse hierarchy theory (RHT) proposed by Ahissar and colleagues (Ahissar and Hochstein, 2004; Ahissar et al., 2009). Both categorization at the superordinate and basic levels would have to be performed in the implicit feed-forward wave of processing, using unconscious high-level representations at the "top" of the visual hierarchy with no need of the reverse hierarchy to low-level representations. It could be that feed-back loops are needed to reach the basic category level when contrasting a target category with a very similar non-target category such as dogs and cats (Bowers and Jones, 2008) or to reach subordinate category level.

Interestingly, when processing the peripheral photograph at the basic level in the dual task, subjects require longer presentation times of the natural scenes than when they are doing superordinate categorization. Note that in such dual task paradigms both the central and peripheral stimuli are masked to ensure that attention cannot be switched from one task to another. Moreover SOA are determined in isolation for the central and the peripheral task and for each participant in order to set task accuracy at 75–80% when participants are performing in the single task condition. As suggested earlier, information processing and especially fine-grained processing, could be highly disrupted by the presence of a mask. Finer, more detailed object representations appear to need more time for information uptake and more integration of perceptual information for selective neurons to fire (Sugase et al., 1999). This need for additional processing could be reflected in the 40- to 65-ms delay found by Macé et al. (2009) between superordinate and basic categorizations for both bird and dog targets.

FAST VISUAL PROCESSING AND "IN CONTEXT" OBJECT REPRESENTATIONS

Whatever the level of categorization of a given object, when tasks are using natural scenes as stimuli, the object is embedded within a background context. In daily life certain objects tend to appear in the same environment, simultaneously, or in close temporal sequence (a hair dryer and a wash basin, a cow and a bush), others on the contrary will seldom appear in the same circumstances (a cow and a wash basin, a hair dryer and a bush). Feature or object selective neurons are part of a large cerebral network in which experience could also shape the strength of mutual interconnections. The visual system is very efficient at extracting stimulus regularities even at a conceptual level. Indeed, subjects can implicitly learn that in a stream of photographs of natural scenes, a bridge is always followed by a bedroom and then by a waterfall (Brady and Oliva, 2008) even though the scene exemplars are always different and even though subjects are performing a task for which scene category is completely irrelevant (responding to a back to-back repeat of the same image). This sort of learning was seen when subjects were asked to report what image would come next in the sequence whether tested with images of the category or with written category word. Thus, implicit learning of statistical regularities can be seen at different level of perceptual abstraction, implying that experience can set-up facilitatory, or inhibitory network connections between populations of selective neurons depending on how frequently (or not) they are coactivated in daily life. Thus even with briefly flashed stimuli, the context in which objects are presented could induce a performance benefit (or cost) depending on whether it is congruent or not with the object category assigned as target (**Figure 6**).

However the strength of contextual influences on object processing and even more the time course of these context/object interactions are still very controversial. In the early 80s, Biederman et al. (1982) showed that the frame of a scene object could interfere with object detection. Detection was less accurate and slower when the object embedded in the scene violated different rules such as

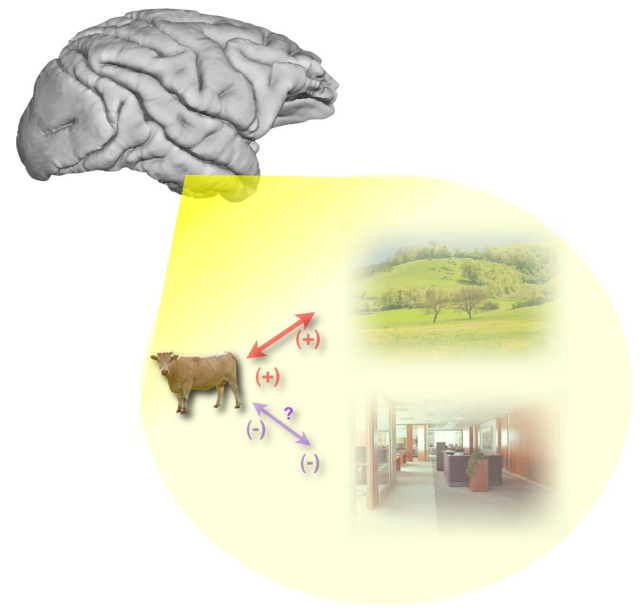

FIGURE 6 | Along the ventral visual pathways, facilitatory/inhibitory connections could develop with expertise between neurons selective for objects or object features that are often/seldom coactivated in daily life (illustration thanks to Maxime Cauchoix).

position, support, size, probability... a set of rules with which painters such as Magritte have been playing with in their artwork. But this hypothesis of a contextual influence on object processing was later challenged by Hollingworth and Henderson (Hollingworth and Henderson, 1998; Henderson and Hollingworth, 1999) who reported that after eliminating guesses and response biases, no advantage was found for the detection of consistent objects over inconsistent ones. They proposed that object identification processes are isolated from knowledge about the world.

The influence of context is often seen as a top down influence. When immersed in a given context, the representations of objects that are likely to appear in such context might be preactivated by expectation (Bar and Ullman, 1996). However in fast visual categorization, subjects are processing a succession of briefly flashed scenes picturing very varied and unrelated scene gists. No predictions can be made about the next photograph that will be presented. Context and embedded objects have to be processed from scratch in an ascending flow of visual processing.

Recent studies (Davenport, 2007; Joubert et al., 2007, 2008) have shown that, in the processing of briefly flashed scenes, when expectations cannot be made, such interference can be demonstrated. In fact, context can interfere with object processing but reciprocally, salient objects can also influence the processing of scene gists. The studies by Joubert et al. (2007, 2008) evaluated such behavioral effects in terms of accuracy and response speed. The presence of a salient object could delay scene categorization (natural vs. urban) by 25–50 ms and even more if the object is incongruent with the scene category, a finding replicated later (Mack and Palmeri, 2010b). Conversely, when objects were embedded in an incongruent context their categorization was less accurate and their processing delayed. The contextual interference on object processing could be seen even in the case of the earliest behavioral responses since MinRT was delayed by 20–30 ms with incongruent object/context associations. Thus, context can modulate object processing even in its very early steps.

To explain such fast interference, a model proposed by Bar's (2004) group suggests that rapid coarse processing of a scene would activate the most likely possible object(s) in a given contextual frame. Activation of the ventral visual areas would be done through the dorsal magnocellular pathway. As shown in **Figure 6**, we would like to propose another model compatible with the processing of a feed-forward processing wave within the ventral pathway. Facilitation would occur when activated populations of neurons selective for visual objects or features are normally used to fire together and have reinforced their interconnections. In other situations activating populations of selective neurons that hardly fire simultaneously, conflicts may arise that can lead to delayed visual processing. In the go–no go categorization task, it was not possible to time the earliest effect of context on object processing because even the earliest responses were affected. We have recently been able to estimate this latency by using a forced choice saccade based categorization task that gives access to an early temporal window (Kirchner and Thorpe, 2006; Crouzet et al., 2010). When this form of saccadic task is used with congruent and incongruent stimulus associations, we found that context could interfere with object processing at around 160 ms (Crouzet et al., 2009).

Using the animal go/no-go visual categorization task, the influence of context on the ascending flow of object information processing has also been demonstrated in monkeys (Fize et al., 2011). Sets of four congruent and incongruent achromatic stimuli (**Figure 7**) were used to prevent all biases possibly induced by natural images. The monkeys' experience is restricted to the thousands of natural images they have been presented with for task performance. Presumably, they must have extracted regular object/context co-occurrences from such experience suggesting that context is used in an implicit way. Although the subject of controversy (Smyth and Shanks, 2008), the implicit use of spatial configuration has been reported in humans performing visual search tasks (Chun and Jiang, 1998, 2003). Chun and Jiang used a task in which certain targets systematically appeared in consistent location within the global configuration. Such targets were detected faster than others embedded in new configurations, but subjects were unable to explicitly discriminate new from repeated stimulus configurations (Chun and Jiang, 1998) or when asked to predict or generate the location of a missing target (Chun and Jiang, 2003). More recently as described earlier, it was also shown that humans could implicitly learn the temporal covariance of semantic categories of natural scenes (Brady and Oliva, 2008) demonstrating that implicit learning can also occur at a conceptual level. After learning, the global features of a scene could be used to modulate the saliency of the different regions of the scene in order for example, to guide visual search to pertinent scene locations (Torralba et al., 2006).

One interesting point in the Fize et al. (2011) paper is the fact that a model only based on scene statistics (Oliva and Torralba, 2001) was at chance level when tested on the stimuli used in this

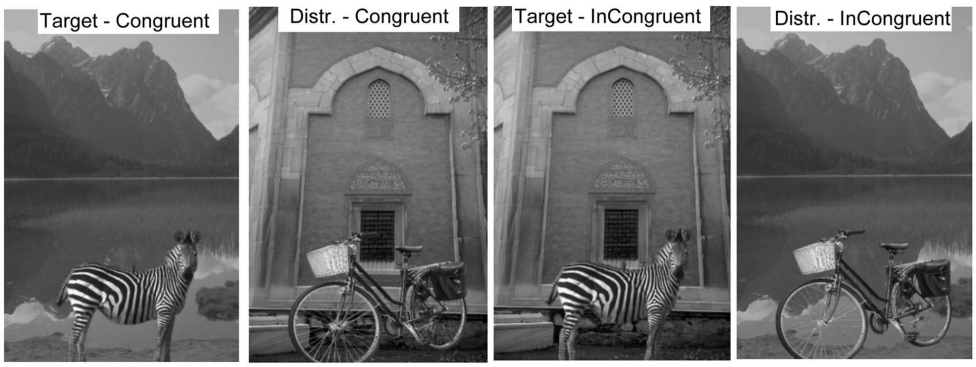

FIGURE 7 | Four stimuli were built using two backgrounds (natural and man made) and two object from different categories (animals and man made objects) resulting in two congruent stimuli (one target and one distracter) and two incongruent stimuli (one target and one distracter). An example of the 192 sets of 4 stimuli that were used in Fize et al. (2011).

study (**Figure 7**). The combination of global scene statistics with the diffusion model of perceptual decision making (Ratcliff, 1978) might have lead to better model performance as suggested by Mack and Palmeri (2010b) for the categorization of scenes containing incongruent objects.

The study by Joubert et al. (2008) was also the first to disentangle the performance impairments due to either stimulus manipulation or object/context congruency. Indeed it showed that pasting an object in another perfectly congruent context is far from being without consequences on object categorization. Even when ensuring that such stimulus manipulation does not violate certain well-established rules (Biederman et al., 1982), it induces a drop of categorization accuracy and response speed. Incidentally, one of the interesting results reported in this study is that performance in fast object categorization is no better for an isolated object when compared to an object presented in its original context. An unsolved question raised by this finding concerns the timing of object segregation. In the ongoing debate about whether object segregation has to precede recognition or whether object recognition influence the figure–ground segregation in briefly presented stimuli (Peterson and Gibson, 1994), our results are more in favor of a late object segregation. Although we found no evidence that scene categorization could be completed faster than object categorization (Joubert et al., 2007), an extreme view would suggest that the visual system may be able to extract the global and ecologically relevant scene primitives needed to describe the spatial lay-out of a scene and conceptual information without the need to segment objects (Oliva and Torralba, 2006, 2007; Torralba et al., 2006; Greene and Oliva, 2009).

CONCLUDING COMMENTS

For the last 15 years, fast go/no-go categorization tasks have thus been useful in describing the characteristics of very early visual processing. They have shown that the first wave of feed-forward processing can allow global coarse object and scene representations at the superordinate category level from achromatic, probably magnocellular, messages. But they also showed the limits of such fast processing when tasks require more detailed object/scene representations. They have demonstrated that early scene processing is largely preattentive and automatic, relying on the regularities of the surrounding world extracted by the visual system through daily experience with no need to require prior object recognition but with processing time courses that are similar for objects and scene gists. The processing of scenes appears massively parallel with facilitatory/inhibitory modulations between object and context processing. The use of the forced choice saccadic task (Kirchner and Thorpe, 2006) will now allow the investigation of an even earlier temporal window and provide new insights into the mechanisms of early visual processing.

Although the processing of briefly presented scenes is required when flipping through a magazine, when zapping from one TV channel to another, and more and more required to follow modern video-clips, the use of briefly flashed stimuli is not usual in our everyday life. Objects and scenes normally persist over time so that information uptake is not restricted by stimulus presentation time. Of course visual processing does not stop after the first wave of early visual processing that has been the focus of the present review. Most daily visual tasks will require the development of attention, the processing of finer details and many feed-back and recurrent processing loops to be solved. Nevertheless the fast visual categorization task has proved very fruitful to investigate the efficiency of the visual system in processing natural scenes under challenging conditions, and at demonstrating the complexity of the tasks that can be completed using only very early, mainly feed-forward and automatic, processing of visual information.

One of the open questions concerns the shaping of object representations by expertise and the temporal advantage offered by expertise (Curby and Gauthier, 2009). Considering an expert as a bird, dog, or car expert, one can predict that through experience, such expert would have developed populations of neurons that are visually selective to very detailed representations of corresponding subordinate categories. The direct prediction that follows is that automatic processing of visual information should be able to reach subordinate level of categorization in expert without any attentional need, whereas novices would be unable to perform the task without attentional resources. Parallel processing of two simultaneously presented scenes would be a very good tool to tackle the limits of preattentive object vision.

REFERENCES

Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* 8, 457–464.

Ahissar, M., Nahum, M., Nelken, I., and Hoschstein, S. (2009). Reverse hierarchies and sensory learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 285–299.

Aubertin, A., Fabre-Thorpe, M., Fabre, N., and Géraud, G. (1999). Processing speed of migraineurs in fast visual categorisation. *C. R. Acad. Sci.* 322, 695–704.

Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci.* 5, 617–629.

Bar, M., and Ullman, S. (1996). Spatial context in recognition. *Perception* 25, 343–352.

Biederman, I. (1972). Perceiving real-world scenes. *Science* 177, 77–80.

Biederman, I., Mezzanotte, R. J., and Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cogn. Psychol.* 14, 143–177

Boucart, M., Naili, F., Despretz, P., Defoort-Delhemme, S., and Fabre-Thorpe, M. (2010). Implicit and explicit object recognition at very large visual eccentricities: no improvement after loss of central vision. *Vis. Cogn.* 18, 839–858.

Bowers, J. S., and Jones, K. W. (2008). Detecting objects is easier than categorizing them. *Q. J. Exp. Psychol. (Colchester)* 61, 552–557.

Brady, T. F., and Oliva, A. (2008). Statistical learning using real-world scenes: extracting categorical regularities without conscious intent. *Psychol. Sci.* 19, 678–685.

Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Brain Res. Rev.* 36, 96–107.

Buswell, G. T. (1935). *How People Look at Pictures: A Study of the Psychology of Perception in Art.* Chicago: University of Chicago Press.

Chun, M. M., and Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn. Psychol.* 36, 28–71.

Chun, M. M., and Jiang, Y. H. (2003). Implicit, long-term spatial contextual memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 224–234.

Crouzet, S. M., Joubert, O. R., Thorpe, S. J., and Fabre-Thorpe, M. (2009). The bear before the forest, but the city before the cars: revealing early object/ background processing. *J. Vis.* 9, 954.

Crouzet, S. M., Kirchner, H., and Thorpe, S. J. (2010). Fast saccades towards faces: face detection in just 100 ms. *J. Vis.* 10, 16.1–17.

Curby, K. M., and Gauthier, I. (2009). The temporal advantage for individuating objects of expertise: perceptual expertise is an early riser. *J. Vis.* 9, 7. 1–13.

Dacey, D. M., and Petersen, M. R. (1992). Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proc. Natl. Acad. Sci. U.S.A.* 89, 9666–9670.

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Mem. Cognit.* 35, 393–401.

De Haan, E. H., Young, A., and Newcombe, F. (1987). Faces interfere with name classification in a prosopagnosic patient. *Cortex* 23, 309–316.

Delorme, A., Richard, G., and Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural images does not rely on colour: a study in monkeys and humans. *Vision Res.* 40, 2187–2200.

Delorme, A., Richard, G., and Fabre-Thorpe, M. (2010). Key visual features for rapid categorization of animals in natural scenes. *Front. Psychol.* 1, 1–13.

Delorme, A., Rousselet, G. A., Macé, M. J.-M., and Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Brain Res. Cogn. Brain Res.* 19, 103–113.

Driver, J., and Vuilleumier, P. (2001). Perceptual awareness and its loss in unilateral neglect and extinction. *Cognition* 79, 39–88.

Elder, J. H., and Velisavljevic, L. (2009). Cue dynamics underlying rapid detection of animals in natural scenes. *J. Vis.* 9, 7.

Fabre-Thorpe, M., Delorme, A., Marlot, C., and Thorpe, S. J. (2001). A limit to the speed of processing in ultra-rapid visual categorisation of novel natural scenes. *J. Cogn. Neurosci.* 13, 171–180.

Fabre-Thorpe, M., Richard, G., and Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *Neuroreport* 9, 303–308.

Fei-Fei, L., Iyer, A., Koch, C., and Perona, P. (2007). What do we perceive in a glance of a real-world scene? *J. Vis.* 7, 1–29.

Ferrera, V. P., Nealey, T. A., and Maunsell, J. H. (1992). Mixed parvocellular and magnocellular geniculate signals in visual area V4. *Nature* 358, 756–761.

Fize, D., Cauchoix, M., and Fabre-Thorpe, M. (2011). Humans and monkeys share visual representations. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7635–7640.

Forti, S., and Humphreys, G. W. (2007). The representation of unseen objects in visual neglect: effects of view and object identity. *Cogn. Neuropsychol.* 24, 661–680.

Gegenfurtner, K. R., and Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Curr. Biol.* 10, 805–808.

Girard, P., Hupe, J.-M., and Bullier, J. (2001). Feedforward and feedback connections between areas v1 and v2 of the monkey have similar rapid conduction velocities. *J. Neurophysiol.* 85, 1328–1331.

Greene, M. R., and Oliva, A. (2009). The briefest of glances: the time course of natural scene understanding. *Psychol. Sci.* 20 464–472.

Grezes, J., Tucker, M., Armony, J., Ellis, R., and Passingham, R. E. (2003). Objects automatically potentiate action: an fMRI study of implicit processing. *Eur. J. Neurosci.* 17, 2735–2740.

Grill-Spector, K., and Kanwisher, N. (2005). As soon as you know it is there, you know what it is. *Psychol. Sci.* 16, 152–160.

Guyonneau, R., Kirchner, H., and Thorpe, S. J. (2006). Animals roll around the clock: the rotation invariance of ultrarapid visual processing. *J. Vis.* 6, 1008–1017.

Henderson, J. M., and Hollingworth, A. (1999). High-level scene perception. *Annu. Rev. Psychol.* 50, 243–271.

Hollingworth, A., and Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *J. Exp. Psychol. Gen.* 127, 398–415.

Jolicoeur, P., Gluck, M. A., and Kosslyn, S. M. (1984). Pictures and names: making the connection. *Cogn. Psychol.* 16, 243–275.

Joubert, O. R., Fize, D., Rousselet, G. A., and Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *J. Vis.* 8, 11.1–18.

Joubert, O. R., Rousselet, G. A, Fize, D., and Fabre-Thorpe, M. (2007). Processing scene context: fast categorization and object interference. *Vision Res.* 47, 3286–3297.

Kaplan, E., and Shapley, R. M. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proc. Natl. Acad. Sci. U.S.A.* 83, 2755–2757.

Kirchner, H., and Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Res.* 46, 1762–1776.

Kosslyn, S. M., Alpert, N. M., and Thompson, W. L. (1995). Identifying objects at different levels of hierarchy: a positron emission tomography study. *Hum. Brain Mapp.* 3, 107–132.

Kruger, K., Donicht, M., Muller-Kusdian, G., Kiefer, W., and Berlucchi, G. (1988). Lesion of areas 17/18/19: effects on the cat's performance in a binary detection task. *Exp. Brain Res.* 72, 510–516.

Li, F. F., VanRullen, R., Koch, C., and Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9596–9601.

Lobue, V., and DeLoache, J. S. (2008). Detecting the snake in the grass: attention to fear-relevant stimuli by adults and young children. *Psychol. Sci.* 19, 284–289.

Macé, M. J.-M. (2006). *Représentations visuelles précoces dans la catégorisation rapide de scènes naturelles chez l'homme et le singe.* Thèse, Université Paul Sabatier – Toulouse 3.

Macé, M. J.-M., Delorme, A. Richard, G., and Fabre-Thorpe, M. (2010). Spotting animals in natural scenes: efficiency of humans and monkeys at very low contrasts. *Anim. Cogn.* 13, 405–418.

Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., and Fabre-Thorpe, M. (2009). Time-course of visual categorizations: you spot the animal faster than the bird. *PLoS ONE* 4, e5927. doi:10.1371/journal.pone.0005927

Macé, M. J.-M., Thorpe, S. J., and Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: how robust at very low contrasts? *Eur. J. Neurosci.* 21, 2007–2018.

Mack, M. L., Gauthier, I., Sadr, J., and Palmeri, T. J. (2008). Object detection and basic level categorization: sometimes you know it is there before you know what it is. *Psychon. Bull. Rev.* 15, 28–35.

Mack, M. L., and Palmeri, T. J. (2010a). Decoupling object detection and categorization. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1067–1079.

Mack, M. L., and Palmeri, T. J. (2010b). Modeling categorization of scenes containing consistent versus inconsistent objects. *J. Vis* 10, 11.1–11.

Martin-Malivel, J., and Fagot, J. (2001). Cross-modal integration and conceptual categorization in baboons. *Behav. Brain Res.* 122, 209–213.

Martinovic, J., Gruber, T., and Muller, M. M. (2008). Coding of visual object features and feature conjunctions in the human brain. *PLoS ONE* 3, e3781. doi:10.1371/journal.pone.0003781

Masquelier, T., Guyonneau, R., and Thorpe, S. J. (2008). Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. *PLoS ONE* 1, e1377. doi:10.1371/journal.pone.0001377

Minamimoto, T., Saunders, R. C., and Richmond, B. J. (2010). Monkeys quickly learn and generalize visual categories without lateral prefrontal cortex. *Neuron* 66, 501–507.

Murphy, G. L., and Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: what a superordinate is good for. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 572–586.

Nealey, T. A., and Maunsell, J. H. (1994). Magnocellular and parvocellular contributions to the responses of neurons in macaque striate cortex. *J. Neurosci.* 14, 2069–2079.

New, J., Cosmides, L., and Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16598–16603.

Nowak, L. G., and Bullier, J. (1997). "The timing of information transfer in the visual system," in *Extrastriate Cortex in Primates*, eds J. Kaas, K. Rocklund, and A. Peters (New York: Plenum), 205–241.

Nowak, L. G., Munk, M. H. J., Girard, P., and Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Vis. Neurosci.* 12, 371–384.

Ohman, A., Flykt, A., and Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *J. Exp. Psychol. Gen.* 130, 466–478.

Oliva, A. (2005). "Gist of the scene," in *Neurobiology of Attention*, eds L. Itti, G. Rees, and J. K. Tsotsos (San Diego, CA: Elsevier), 251–256.

Oliva, A., and Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cogn. Psychol.* 41, 176–210.

Oliva, A., and Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42, 145–175.

Oliva, A., and Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Prog. Brain Res.* 155, 23–36.

Oliva, A., and Torralba, A. (2007). The role of context in object recognition. *Trends Cogn. Sci.* 11, 520–527.

Palmeri, T. J., and Gauthier, I. (2004). Visual object understanding. *Nat. Rev. Neurosci.* 5, 291–303.

Peterson, M. A., and Gibson, B. S. (1994). Must figure ground organization precede object recognition? *Psychol. Sci.* 5, 253–259.

Poncet, M., Reddy, L., and Fabre-Thorpe, M. (2011). Animal or dog? Vehicle or car? The answer lies in more information not attention. ECVP Toulouse-France 28 août-1er Sept. *Perception suppl.* 40, 33.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learn. Mem.* 2, 509–522.

Potter, M. C., and Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *J. Exp. Psychol.* 81, 10–15.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108.

Reddy, L., Reddy, L., and Koch, C. (2006). Face identification in the near-absence of focal attention. *Vision Res.* 46, 2336–2343.

Reddy, L., Wilken, P., and Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *J. Vis.* 4, 106–117.

Roberts, W. A., and Mazmanian, D. S. (1988). Concept learning at different levels of abstraction by pigeons, monkeys, and people. *J. Exp. Psychol. Anim. Behav. Process.* 14, 247–260.

Rogers, T. T., and Patterson, K. (2007). Object categorization: reversals and explanations of the basic-level advantage. *J. Exp. Psychol. Gen.* 136, 451–469.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cogn. Psychol.* 8, 382–439.

Rousselet, G. A., Fabre-Thorpe, M., and Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nat. Neurosci.* 5, 629–630.

Rousselet, G. A., Macé, M. J.-M., and Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *J. Vis.* 3, 440–456.

Rousselet, G. A., Thorpe, S. J., and Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision Res.* 44, 877–894.

Schyns, P. G., and Oliva, A. (1994). From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychol. Sci.* 5, 195–200.

Sherman, S. M. (1985). Functional organization of the W-, X- and Y-cell pathways in the cat: a review and hypothesis. *Prog. Psychobiol. Physiol. Psychol.* 11, 233–314.

Silveira, L. C., and Perry, V. H. (1991). The topography of magnocellular projecting ganglion cells (M-ganglion cells) in the primate retina. *Neuroscience* 40, 217–237.

Smyth, A. C., and Shanks, D. R. (2008). Awareness in contextual cuing with extended and concurrent explicit tests. *Mem. Cognit.* 36, 403–415.

Spence, I., Wong, P., Rusan, M., and Rastegar, N. (2006). How color enhances visual memory for natural scenes. *Psychol. Sci.* 17, 1–6.

Strasburger, H., and Rentschler, I. (1996). Contrast-dependent dissociation of visual recognition and detection fields. *Eur. J. Neurosci.* 8, 1787–1791.

Sugase, Y., Yamane, S., Ueno, S., and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400, 869–873.

Tanaka, J. W., and Curran, T. (2001). A neural basis for expert object recognition. *Psychol. Sci.* 12, 43–47.

Tanaka, J. W., and Taylor, M. (1991). Object categories and expertise: is the basic level in the eye of the beholder? *Cogn. Psychol.* 23, 457–482.

Thorpe, S. J., and Fabre-Thorpe, M. (2001). Neuroscience. Seeking categories in the brain. *Science* 291, 260–263.

Thorpe, S. J., and Fabre-Thorpe, M. (2003). "Fast visual processing and its implications," in *The Handbook of Brain Theory and Neural Networks*, 2nd Edn, ed. M. Arbib (Cambridge: MIT press), 441–444.

Thorpe, S. J., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522.

Thorpe, S. J., Gegenfurtner, K., Fabre-Thorpe, M., and Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *Eur. J. Neurosci.* 14, 869–876.

Torralba, A., Oliva, A., Castelhano, M. S., and Henderson, J. M. (2006). Contextual guidance of eye-movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* 113, 766–786.

VanRullen, R., Reddy, L., and Koch, C. (2004). Visual search and dual tasks reveal two distinct attentional resources. *J. Cogn. Neurosci.* 16, 4–14.

VanRullen, R., and Thorpe, S. J. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception* 30, 655–668.

VanRullen, R., and Thorpe, S. J. (2001b). The time course of visual processing: from early perception to decision-making. *J. Cogn. Neurosci.* 13, 454–461.

Vidyasagar, T. R. (1999). A neuronal model of attentional spotlight: parietal guiding the temporal. *Brain Res. Brain Res. Rev.* 30, 66–76.

Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. *Eur. J. Neurosci.* 11, 1223–1238.

Wichmann, F. A., Sharpe, L. T., and Gegenfurtner, K. R. (2002). The contributions of color to recognition memory for natural scenes. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 509–520.

Yao, A. Y., and Einhauser, W. (2008). Color aids late but not early stages of rapid natural scene recognition. *J. Vis.* 8, 11–13.

Yarbus, A. L. (1961). Eye movements during the examination of complicated objects. *Biofizika* 6, 52–56.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 June 2011; accepted: 05 September 2011; published online: 03 October 2011.

*Citation: Fabre-Thorpe M (2011) The characteristics and limits of rapid visual categorization. Front. Psychology **2**:243. doi: 10.3389/fpsyg.2011.00243*

This article was submitted to Frontiers in Perception Science, a specialty of Frontiers in Psychology.

Copyright © 2011 Fabre-Thorpe. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.