UvA-DARE (Digital Academic Repository)

Conscious and unconscious vision

Fahrenfort, J.J.

[Link to publication](#)

Chapter 5.

Object classification in the absence of visual awareness and figure-ground segregation

Abstract

It is well known that neurons in the temporal lobe classify objects, such as faces, and it is generally assumed that the activity of such neurons is necessary for conscious awareness of these objects. However, object categorization may also occur unconsciously, as has been shown by the selective activation of object selective neurons by masked objects. So what distinguishes conscious from unconscious object recognition? We constructed schematic images containing objects such as faces and houses while keeping local retinal stimulation between conditions identical. Using a dichoptic fusion paradigm, we manipulated stimulus visibility such that objects were either visible or not visible. Confirming earlier results, we found that both consciously perceived and non-perceived objects result in category specific BOLD activation, even if they are task irrelevant and non-attended. Critically however, we show that objects that are consciously seen show a distinct neural signature of figure-ground segregation in early and midlevel visual areas, which is completely absent when objects are not seen. Although counterintuitive, this implies that consciousness is more intimately related to processes of figure-ground segregation and perceptual organization than to object categorization. We propose that figure-ground segregation is a prerequisite for visual awareness, and that both phenomena share part of their neural correlate, which is recurrent processing within visual cortex.

Introduction

Some regions in the visual cortex are particularly sensitive to processing specific object categories. Amongst the first experiments showing this were a study identifying an area in the fusiform gyrus engaged in processing faces (Kanwisher, et al., 1997) and a study showing a region in the parahippocampus engaged in processing houses or scenes containing a spatial layout (Epstein & Kanwisher, 1998), both of which extended the observation that a large region termed the lateral occipital complex (LOC) responds more strongly to objects than to scrambled objects or other visual patterns (Malach, et al., 1995).

It has also been shown that the brain is able to perform such object categorization even when objects cannot be consciously perceived (Kouider, et al., 2008; G. Kovacs, et al., 1995; Marois, et al., 2004; Moutoussis & Zeki, 2002). This implies that object categorization is not the same as object perception, although it is still unclear what cortical mechanisms distinguish the two. So what does it take to go from object classification to actually seeing something like a face? A number of studies suggest that object categorization is highly distributed (e.g. Haxby, et al., 2001; Schwarzlose, Swisher, Dang, & Kanwisher, 2008). However, the objects in these studies can be readily perceived and are consciously categorized. As many studies have implicated recurrent processing in conscious visual perception (e.g. Boehler, Schoenfeld, Heinze, & Hopf, 2008; Haynes, et al., 2005; Silvanto, et al., 2005; Super, et al., 2001), it may be that distributed category representations emerge only as a result of conscious visual perception. There is evidence that object categorization can also proceed in a highly automated fashion, based on feedforward processes alone (G. Kovacs, et al., 1995; Thorpe, et al., 1996; VanRullen & Koch, 2003). Although speculative, we hypothesize from these results that conscious object perception may be tied to distributed representations resulting from recurrent processing, whereas unconscious object categorization is tied to fleeting focal representations resulting from feedforward processing.

This also has bearing on how object categorization is related to processes of figure-ground segregation. Traditionally, the logical sequence of events was thought to be that objects were first

segregated from their background before they could be classified (e.g. Rubin, 1958). As figure-ground segregation has repeatedly been associated with recurrent processing (e.g. Lamme, 1995) and in light of the hypothesis that object categorization is an unconscious feedforward process, we postulate the reverse, that object categorization precedes figure-ground segregation and as such may proceed independently from it. In our proposed scheme of events, both unseen and seen objects are initially categorized by category selective neurons, whereas only objects that are subsequently segregated from their background are consciously perceived. Although counterintuitive, this would imply that consciousness is more intimately related to processes of figure-ground segregation and perceptual organization than to object categorization.

To test these hypotheses we used a dichoptic fusion paradigm in which stimuli were either visible or invisible (see Figure 5.1a). Using classical contrasts specific to object categorization and figure-ground segregation we were able to directly test the relationship between object categorization, figure-ground segregation and visual awareness.

Methods

Participants

Eighteen subjects participated in the experiment for a monetary reward. All subjects provided written informed consent. All procedures were approved by the ethical committee of the University of Amsterdam. Two subjects were not included in the analyses: one subject's data was lost due to human error, the other subject showed excessive movement during scanning. Sixteen subjects (3 male, 2 left handed) with normal vision and an average age of 21.1 years (± 3.4) were included in data analysis.

Stimuli

Stimuli were presented at 800 × 600 resolution with a projected size measuring 16.9° × 12.7° visual angle. They were composed of a 28 by
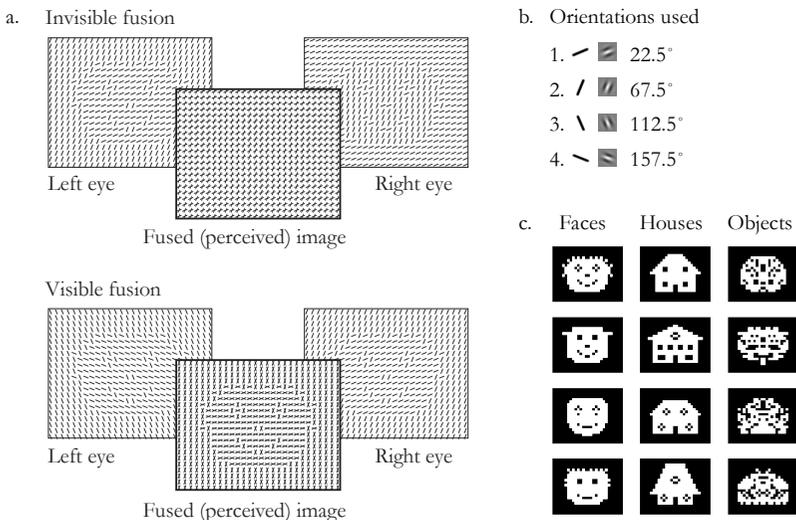
Figure 5.1 *(a) Example of invisible and visible dichoptic fusion using schematic line elements. In either eye, objects were created in the image by using a different orientation for the Gabor elements in the foreground (depicting a face in this example) and the background. There was always a 45° orientation difference between foreground and background. For invisible fusion (top panel), objects were created by using the same orientation between elements belonging to the foreground of one eye and elements in the background of the other eye. This way, images in either eye contain an object whereas the percept of the dichoptically fused image consists of a homogenous screen of crossed Gabor elements. For visible fusion (bottom panel) objects were created by using a different orientation for the elements in the foreground of one eye and elements in the background of the other eye, resulting in differently crossed Gabor elements for foreground and background. (b) Schematic lines depicting Gabor elements, as well as the actual Gabor elements used in the experiment. Four orientations were used: 22.5°, 67.5°, 112.5° and 157.5°. (c) Schematic depiction of all four instances of three of the four stimulus categories: faces, houses and nonsense-objects (homogenous screens are not shown). Black represents one orientation, white the other orientation.*

22 matrix of oriented 2-dimensional Gabor functions, oriented elements with luminance features that roughly model the receptive field structure of V1 simple cells (period $\approx$ 0.58°, $\sigma \approx$ 0.10°, schematically represented by lines in the left and right eye of Figure 5.1a). By using a different orientation for the Gabor elements representing the figure (the foreground) and for elements representing the background, objects could be constructed in the images. Four

orientations were used: 22.5°, 67.5°, 112.5° and 157.5° (see schematic lines and actual Gabor elements in Figure 5.1b). An image containing an object would always be composed of two orientations with a 45° difference between them (for examples see left and right eye of Figure 5.1a). Four stimulus categories were created in this way: faces, houses, nonsense objects and homogenous screens using all possible orientations (see schematic representation in Figure 5.1c, white is one orientation, black the other orientation; only stimulus categories containing an object are shown, homogenous screens are omitted in this schematic). There were four instances of each of the object categories (see Figure 5.1c). All objects consisted of the same number of oriented elements.

Visible versus invisible dichoptic fusion

Stimuli containing different orientations of Gabor elements were presented to the left and the right eye. When different images are presented to the left and right eye for a short duration, they are fused to a single percept, a phenomenon known as dichoptic fusion (see Kolb & Braun, 1995; Moutoussis & Zeki, 2002; Zipser, et al., 1996). The fusion phenomenon was exploited to create conditions under which objects in the images were either visible or not visible, while keeping physical stimulus characteristics locally identical.

Dichoptically invisible objects were created by using the same orientation between the elements belonging to the object (the foreground) of one eye and the elements in the background of the other eye, while maintaining a 45° difference between foreground and background within each eye. This way, images in either eye contain an object whereas the percept of the dichoptically fused image consists of a homogenous screen of crossed Gabor elements (see Figure 5.1a top). Dichoptically visible objects were created in the same way, but using a different orientation for the elements in the foreground of one eye and the elements in the background of the other eye, resulting in differently crossed elements for foreground and background (see Figure 5.1a bottom).

This was done for all objects and homogenous screens, thus creating 8 conditions:
dichoptically fused visible faces, houses, nonsense objects and homogenous screens, and dichoptically fused invisible faces, house,

nonsense objects and homogenous screens. For brevity, dichoptically fused visible and dichoptically fused invisible conditions are henceforth referred to simply as visible and invisible conditions.

Stimulus sequence

The sequence of events in a trial can be found in Figure 5.2a. A red fixation dot was present in the middle of the screen throughout the experiment. Each trial started with 167 ms in which a grey screen was shown. After this, the stimulus sequence was presented four times in succession. Each sequence began with a 50 ms mask, followed by an isoluminant grey screen for 67 ms, the stimulus screen for 83 ms and ending with another isoluminant grey screen for 50 ms. The (repeated) short stimulus presentation of 83 ms was used to aid dichoptic fusion. The masking screens (acting both as forward and backward masks) consisted of a 28 by 22 sized matrix of Gabor elements of which half were randomly rotated to an orientation of 0° and the other half were rotated to 90°. If the orientation of an element was at 0° in the left eye it would be at 90° in the right eye and vice versa, resulting in a homogenously fused percept consisting of plus-like Gabor elements.
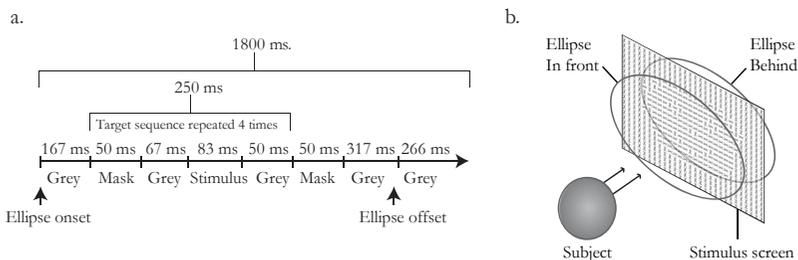
Figure 5.2 *(a) Schematic depiction of the timeline of a trial. (b) 3D representation of a subject performing the ellipse "in front" / "behind" task (both "in front" and "behind" are shown for illustrative purposes). An illusion of dimensionality was created of an ellipse hovering "in front" or "behind" the stimulus screen by offsetting ellipse location horizontally from center in different directions for the left and the right eye.*

On each trial, the sequence of four stimulus presentations always contained one stimulus category. Half of the trials contained a sequence of four of the same instances of its stimulus category (Same trials); the other half of the trials contained four different instances of its stimulus category (Different trials). These were originally intended to look at adaptation effects, but explorative analyses showed that statistical power proved insufficient for useful analysis of such adaptation effects. Since the Same-Different manipulation had no bearing on the rest of the experiment, Same-Different trials were pooled in all subsequent analyses and are mentioned no further.

In addition, during the first 1533 ms of a trial, an elliptic circle of size 15.45° × 11.76° visual angle overlaying the stimulus sequence was presented to each eye (large enough to encompass faces/houses/objects in the stimulus sequence). By displacing each ellipse 7 pixels from the fixation center in opposite direction for each eye, a retinal disparity was created. This disparity created an illusion of dimensionality, in which the perceived ellipse either seemed to hover in front or behind the stimulus screen, but only when properly focusing with both eyes. A three dimensional impression of what this looks like can be found in Figure 5.2b. On half of the trials, the disparity was such that the circle seemed to hover in front of the stimulus screen, on the other half of the trials the circle appeared to hover behind the stimulus screen, although the disparity was small enough to make this difficult to see when not paying attention, and impossible to see when viewing the stimuli monocularly. During the experiment, subjects were instructed to indicate whether the circle was hovering 'in front' or 'behind' the stimulus sequence. This was done to induce proper convergence for stereoscopic viewing, a necessary condition for dichoptic fusion to occur. As soon as target stimulus sequence began, subjects were able to give a response on the ellipse task, until the next trial would begin. Trials were presented in immediate succession, although sometimes empty dummy trials lasting 1.8 or 3.6 seconds were presented in between (see fMRI acquisition).

The orientations of the Gabor elements in the stimuli and all other elements in the stimulus sequence were balanced out in such a way that local physical stimulation was on average the same for faces, houses, nonsense-objects and homogenous screens within the visible

and invisible conditions. This allowed us to make contrasts between conditions that are sensitive to higher order differences in perceptual organization, while largely shutting out the influence of local physical stimulation on cortical processing (such as the influence of the Gabor elements themselves on orientation selective neurons in early visual areas, Hubel & Wiesel, 1962, 1968; Lamme, 1995; Zipser, et al., 1996).

Tasks

On each trial during the experiment, subjects had to indicate whether an ellipse was hovering 'in front' or 'behind' the stimulus sequence using a two-button response panel in their right hand (see under Stimulus Sequence above for a description of ellipse presentation). The reason for using this task was threefold: [1] using this task forced subjects to view the stimuli stereoscopically, which aids dichoptic fusion [2] using an attentionally demanding task that does not focus on the stimuli themselves, makes it less likely that differences between the visible and invisible conditions are caused by differences in attention or task relevance and [3] it is less likely that any correlates of visibility found in this study are related to the act of report about the stimulus.

To determine whether subjects were successfully fusing stimuli in the experiment, a control task was performed after scanning (but still inside the scanner), in which subjects had to make two consecutive responses on each trial: [1] indicate whether the circle was hovering 'behind' or 'in front' of the stimulus sequence, and subsequently [2] indicate the object category in the stimulus sequence (face, house, nonsense object or homogenous screen). If subjects were successfully fusing the objects, one would predict that invisible faces, houses and nonsense objects would be classified as homogenous screens, whereas visible stimuli would be correctly categorized. Responses were made on a two-button response panel in their right hand (in front / behind) and a four-button response panel in their left hand (faces, houses, nonsense objects and homogenous screens). This control task consisted of 32 trials per condition in random order, a total of 256 trials.

Apparatus for stereo presentation

To be able to present different images to the left and the right eye, all visual stimulation was presented using two stacked beamers projecting at exactly the same location on a projector screen in the scanner room. Using polarizing filters, the upper beamer projected light of opposite polarity compared to the light of the bottom beamer. Subjects viewed the screen via a mirror through polarizing glasses. The glass of the left eye only allowed light of the upper beamer to pass through, the glass of the right eye only allowed light of the lower beamer to pass through.

fMRI acquisition, preprocessing and analysis

Data was collected on a Philips 3T Intera scanner. All subjects took part in one scanning session, which always started with the acquisition of a high resolution three dimensional anatomical image of the head using a T1 image sequence lasting about 6 minutes. Subsequently, three runs of functional data were collected, except for three subjects from whom four runs were acquired. A functional run started and ended with 16 seconds rest and lasted approximately 12 minutes. The functional recordings were acquired using a T*-weighted sequence (TR 2301 ms, TE 28 ms, 35 slices, slice thickness 3 mm; slice gap 0.3 mm, FOV 220 × 220 mm). Each run contained 256 target trials, in which each of the eight stimulus categories were presented 32 times in random order. In addition, 100 dummy trials containing only an isoluminant grey screen intended to improve deconvolution and counteract BOLD saturation in visual areas were randomly interspersed between these trials. Of these dummy trials, 20 lasted 3.6 seconds, and 80 lasted 1.8 seconds.

Preprocessing and data analysis was performed using FSL (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl) and Matlab (The Mathworks Inc., Natick, MA, USA). Using FEAT (fMRI Expert Analysis Tool, v5.92) fMRI images were motion corrected using MCFLIRT, slice-time aligned, brain areas were extracted using BET, the data were spatially smoothed using a Gaussian kernel of 5 mm and high-pass temporally filtered using a Gaussian envelope (σ = 35 s).

All functional data were aligned to the structural image of the subject. Using the structural image, the data of each subject was transformed to the standard space of the Montreal Neurological Institute (MNI) using FLIRT. Subsequently all conditions and relevant contrasts were specified in a General Linear Model (GLM) and all runs were analyzed using this GLM. Runs were pooled on a per subject basis using a fixed effects model. Subsequently, a mixed effects group analysis was done (FMRIB's FLAME stage 1) in which relevant lower level contrasts were combined. These were used to help determining the regions of interest. From these regions of interest, parameter estimates from the GLM were transformed into estimates representing percent signal change and exported for each subject and each condition to perform final analyses.

Regions of interest (ROI's)

We looked at four regions of interest: [1] the FFA (Fusiform Face Area), [2] a region in the subiculum of the hippocampus that responded to the contrast normally used to locate the PPA (Parahippocampal Place Area), [3] the LOC (Lateral Occipital Cortex), which was subdivided into a superior and an inferior part and [4] V1 and V2. The exact regions for which we compared visible and invisible stimuli were chosen because of their theoretical interest, and defined on the basis of a combination of anatomical and functional characteristics. All functional contrasts used in defining the ROI's (FFA, PPA, LOC and V1/V2) were created by pooling visible and invisible stimuli, so as not to bias the results in any direction towards the processing of visible or invisible stimuli.

The left and right FFA ROI's were defined using a classical contrast used to locate the FFA (Kanwisher, et al., 1997): voxels responding more strongly to Faces than to houses and other objects. We included voxels that exceeded an uncorrected threshold of Z > 2.3 within the area of the Temporal Occipital Fusiform Cortex as defined by the Harvard-Oxford Cortical Structural Atlas in MNI space. This resulted in two well-defined clusters corresponding to the FFA as described in the literature, with a larger cluster for the right hemisphere than for the left hemisphere (see Kanwisher, et al., 1997). Voxels in the ROI's were weighted more strongly as they responded

more strongly to faces than to houses and nonsense objects, as this is the contrast that classically defines the FFA.

Originally, we had intended to look at activity in the PPA (Parahippocampal Place Area), which has been implicated in perceiving the local visual environment, such as represented by rooms, scenes and houses (Epstein & Kanwisher, 1998). However, the houses > faces & objects contrast which is typically used to identify the PPA, did not result in convincing clusters representing typical PPA activation in the parahippocampus. However, a symmetrical bilateral cluster of voxels did show up slightly more dorsal, in the subiculum of the hippocampus itself, as structure that has been implicated in, amongst other things, the encoding of novel complex pictures and other information (e.g. Clark, Broadbent, Zola, & Squire, 2002; Stern, et al., 1996). The reason for the classical PPA contrast not showing up in the parahippocampus may well have something to do with the fact that the PPA is specifically sensitive to representations containing the layout of the local environment, such as scenes and rooms and much less to objects without a three-dimensional spatial context (even if they contain a spatial representation, Epstein & Kanwisher, 1998). Our iconic reproduction of a house using Gabor elements may simply have not had enough flavor of the local spatial environment. However, we did include the slightly more dorsal bilateral cluster in the hippocampus. It was defined by voxels exceeding an uncorrected threshold of $Z > 2$ in the subiculum of the hippocampus as defined by the Juelich Histological Atlas in MNI space. Voxels were weighted more heavily as they responded more strongly to houses than to faces and nonsense objects.

The LOC regions of interest (inferior and superior LOC separately) were chosen because of the implication of the LOC in human object recognition (Grill-Spector, Kourtzi, & Kanwisher, 2001). The regions were defined as those clusters of voxels responding more to stimuli containing a figure-ground relationship (faces, houses and nonsense objects) than to homogenous screens (thresholded by a $Z > 2.3$ and a corrected cluster significance threshold of $p = .05$) lying either in the inferior LOC or superior LOC as defined by the probabilistic Harvard-Oxford Cortical Structural Atlas in MNI space. Voxels in the ROI's were weighted

more strongly as they had a bigger probability of belonging to either the superior or to the inferior LOC as defined by the atlas.

Finally, V1 and V2, were chosen because of their traditional implication in low level vision and the more recent implication of V1 in figure-ground segregation and visual awareness through recurrent processing (Lamme, 1995; Lamme, et al., 2000). The ROI's were defined as those clusters of voxels responding more strongly to stimuli containing a figure-ground relationship (faces, houses and nonsense objects) than to homogenous screens (thresholded by a Z > 2.3 and a corrected cluster significance threshold of p = .05) lying either in V1 or in V2, as defined by the probabilistic Juelich Histological Atlas in MNI space. Again, voxels in these ROI's were weighted according to the probability that they belonged to either V1 or V2.

Results

Behavior

In a behavioral control task performed directly after scanning (but still inside the scanner), subjects had to make two responses: [1] the ellipse localization response which they also performed during scanning (see Methods), and [2] a stimulus categorization response. The stimulus categorization task was intended as confirmation that subjects were fusing well enough to render stimuli invisible in the invisible condition. The average response frequency for each of the stimulus categories can be found in Figure 5.3, separately for invisible (left panel) and for visible trials (right panel). As can be seen in Figure 5.3 (left panel), almost all of the invisible trials were categorized as homogenous screens, thus indicating that fusion and our visible-invisible manipulation was successful. One sample *t*-tests against zero confirmed that none of the invisible categories (except homogenous screens) were identified above chance level. Figure 5.3 (right panel) shows that all of the categories in the visible condition were identified well above chance level.

During scanning, subjects only had to perform the ellipse localization ('in front'/'behind') task. Despite task difficulty, the average proportion correct was .89 (SEM .02), indicating that subjects were viewing the stimuli stereoscopically. The percentage correct did

not differ between 'in front' or 'behind' trials [$F_{1,15}$ = .35, p = .57], or between visible and invisible trials [$F_{1,15}$ = .54, p = .47], as determined by a repeated measures test. On the control task, subjects scored .91 (SEM .03) on the ellipse localization task. Subjects did not score differently on ellipse localization during scanning and ellipse localization during the control task [t_{15} = 1.28, p = .22], confirming that stereoscopic viewing conditions between the control task and the experiment itself were highly comparable.
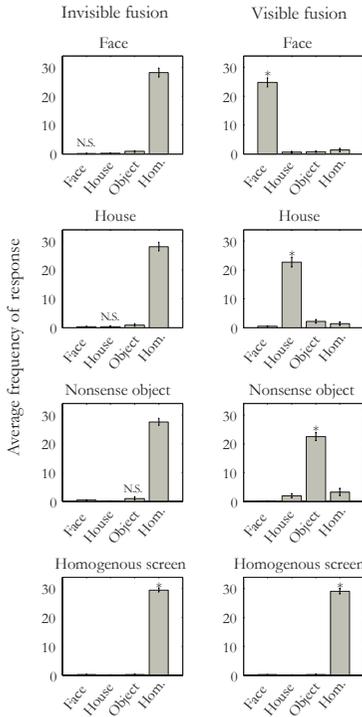
Figure 5.3 *Behavioral response frequency data for (a) invisible conditions and (b) visible conditions. Each graph represents a stimulus condition. Mean response frequency ($\pm$1 s.e.m.) is given for all possible responses. For each condition a one sample t-test against 0 was performed to determine whether that category was identified above chance level [N.S. = p > .05; * = p < 10^{-15}].*

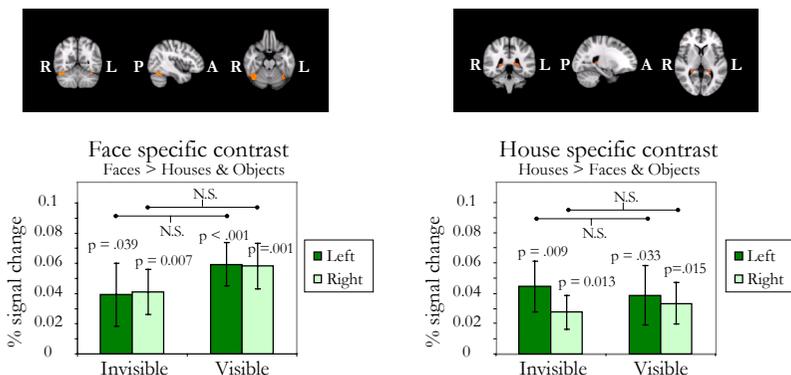## *Object specific activation for face and house stimuli*

We aimed to reveal whether face and house specific activity could be found for visible and invisible stimuli. Parameter estimates for faces were tested against parameter estimates for houses & objects using one-tailed paired t-tests, separately for visible and invisible stimuli, and separately for the left and right FFA (see Figure 5.4a). Both

visible and invisible faces caused significantly more activation in both the left and right FFA when compared to houses & objects, thus showing face specific activation for both visible and invisible faces. Moreover, as can bee seen in the bottom panel of Figure 5.4a, there was no significant difference between the visible and invisible FFA contrast for either the left or the right FFA (see figure and legend for *t*- and *p*-values).

a. FFA ROI's in fusiform cortex   b. ROI's in hippocampal area

Figure 5.4 *(a) face and (b) house specific activation for visible and invisible stimuli.*
(a, top panel) Left and right ROI's showing face selective voxels in orange. ROI's were determined using the faces > houses & objects contrast (see methods). Voxels with higher Z-values have lighter colors. (a, bottom panel) Mean % signal change (±1 s.e.m.) for the faces > houses & objects contrast, separately for visible and invisible stimuli in the left and right FFA. Paired t-tests confirmed that both invisible [left FFA: t_{15} = 2.27, p = .039, right FFA: t_{15} = 2.78, p = .007] and visible stimuli [left FFA: t_{15} = 4.50, p < .001, right FFA: t_{15} = 4.15, p = .001] showed above chance % signal change. Moreover, paired t-tests between visible and invisible conditions showed that visible stimuli did not show significantly more activation than invisible stimuli in these areas [left FFA: t_{15} = 1.29, p = .215, right FFA: t_{15} = 1.29, p = .218].
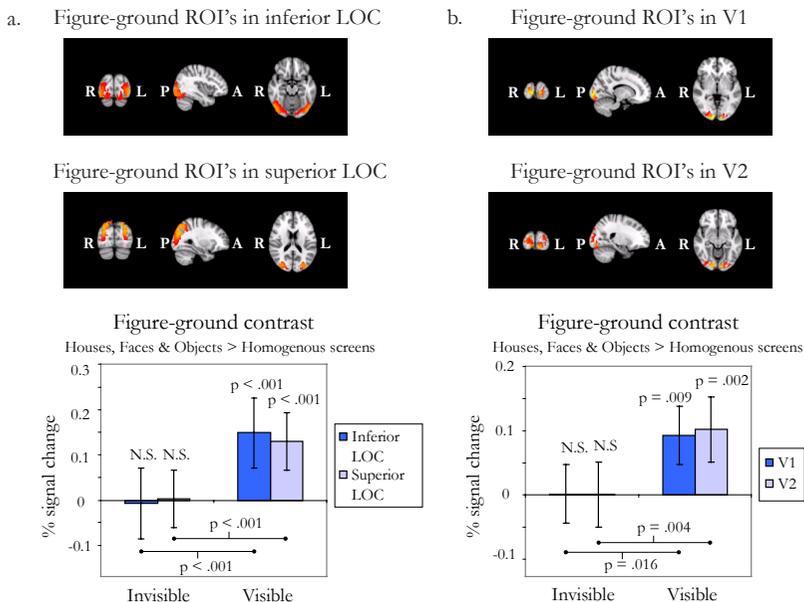*(b, top panel) Left and right ROI's showing house selective voxels in orange. ROI's were determined with the houses > faces & objects contrast (see methods). Voxels with higher Z-values have lighter colors. (b, bottom panel) Mean % signal change (±1 s.e.m.) for the houses > faces & objects contrast, separately for visible and invisible stimuli in the left and right ROI. Paired t-tests confirmed that both invisible [left ROI: t_{15} = 2.99, p = .009, right ROI: t_{15} = 2.83, p = .013] and visible stimuli [left ROI: t_{15} = 2.35, p < .033, right ROI: t_{15} = 2.76, p = .015] showed above chance % signal change. Moreover, paired t-tests between visible and invisible conditions showed that visible stimuli did not show significantly more activation than invisible stimuli [left ROI: t_{15} = .83, p = .415, right ROI: t_{15} = .91, p = .38].*

One-tailed paired *t*-tests of houses against faces & other objects revealed house specific activation, both for visible and invisible houses, and both for the left and the right hemispheric ROI (see Figure 5.4b). There was no significant difference between the visible and invisible contrasts for the ROI in the left or for the ROI in the right hemisphere (see bottom panel of figure 4b and legend for *t*- and *p*-values). In summary, it appeared that both houses and faces evoked category specific activation of high level areas, regardless of their visibility, confirming earlier results using a similar dichoptic fusion paradigm (Moutoussis & Zeki, 2002).

Figure-ground activity in LOC and early visual areas for visible, but not for invisible stimuli

Subsequently, we aimed to reveal activity related to figure-ground segregation, by contrasting stimuli containing a surface with those that do not contain a surface (houses, faces & other objects > homogenous screens). From the voxels in this contrast we created ROI's using anatomical criteria (see Methods section), resulting in four ROI's: [1] inferior LOC, [2] superior LOC, [3] V1 and [4] V2 (see Figure 5.5a and b, top panels). Note that all figure-ground activity was covered by these four occipital ROI's, there was no additional activity in more frontal or other regions of the cortex.

Activity caused by faces, houses and objects was compared to activity caused by homogenous screens using paired t-tests, separately for visible and invisible stimuli. This resulted in significant activation for visible stimuli in inferior LOC, superior LOC (Figure 5.5a, bottom panel) and V1 and V2 (Figure 5.5b, bottom panel). Invisible stimuli however, showed no significant activation whatsoever for this contrast in any of these areas. Moreover, when comparing visible and invisible contrasts directly, visible contrasts show significantly more activity than invisible contrasts in all of these areas (see bottom panels of Figure 5.5a and 5.5b, legend for *t*- and *p*-values). The same was true for all object categories separately. When doing separate *t*-tests, invisible faces, houses and objects did not show more activity than homogenous screens in these areas, whereas visible faces, houses and objects did.

a.   Figure-ground ROI's in inferior LOC

Figure-ground ROI's in superior LOC

b.   Figure-ground ROI's in V1

Figure-ground ROI's in V2

Figure 5.5 *Figure-ground activation in (a) inferior and superior LOC and in (b) V1 and V2.*
(a, top panel) Inferior and superior LOC ROI's in orange containing those voxels that are selective for stimuli containing a figure-ground relationship (see methods). Voxels with a higher probability of belonging to inferior or superior LOC are indicated by lighter colors. (a, bottom panel) Mean % signal change (± 1 s.e.m.) for the faces, houses & objects > homogenous screens contrast, separately for visible and invisible stimuli in inferior and superior LOC. Paired t-tests confirmed that invisible did not show above chance % signal change [LOC inferior: $t_{15} = .96$, $p = .352$, LOC superior: $t_{15} = .77$, $p = .454$], whereas visible stimuli did [LOC inferior: $t_{15} = 6.51$, $p < .001$, LOC superior: $t_{15} = 6.19$, $p < .001$]. Moreover, paired t-tests between visible and invisible conditions showed that visible stimuli showed significantly more activation than invisible stimuli [LOC inferior: $t_{15} = 5.41$, $p < .001$, LOC superior: $t_{15} = 4.42$, $p < .001$].
(b, top panel) V1 and V2 ROI's in orange containing those voxels that are selective for stimuli containing a figure-ground relationship (see methods). Voxels with a higher probability of belonging to V1 or V2 respectively are shown in lighter colors. (b, bottom panel) Mean % signal change (± 1 s.e.m.) for the faces, houses & objects > homogenous screens contrast, separately for visible and invisible stimuli in V1 and V2. Paired t-tests confirmed that invisible did not show above chance % signal change [V1: $t_{15} = .73$, $p = .478$, V2: $t_{15} = .71$, $p = .490$], whereas visible stimuli did [V1: $t_{15} = 3.03$, $p = .009$, V2: $t_{15} = 3.74$, $p = .002$]. Moreover, paired t-tests between visible and invisible conditions showed that visible stimuli showed significantly more activation than invisible stimuli [V1: $t_{15} = 2.72$, $p = .016$, V2: $t_{15} = 3.44$, $p < .004$].

This was true for all areas, except for the visible faces versus homogenous t-test in V1 (p = .075), possibly due to a lack of statistical power (see Table 5.1 for *t*- and *p*-values for all *t*-tests in all areas). Inferior LOC, superior LOC, V1 and V2 all seem to be engaged in figure-ground segregation, but only when the segregated object is consciously perceived.

Table 5.1 *Figure ground contrast for individual stimulus categories. t- and p-values for the individual t-tests between the object categories and homogenous screens, separately for visible and invisible conditions in inferior and superior LOC as well as V1 and V2. All visible categories except faces in V1 show significant p-values as indicated by an asterisk (*), whereas none of the invisible categories do so.*

		LOC inferior		LOC superior		V1		V2	
		t-value	p-value	t-value	p-value	t-value	p-value	t-value	p-value
invisible	faces > homogenous	0.95	0.356	1.56	0.140	1.07	0.300	1.06	0.305
	houses > homogenous	1.25	0.230	1.03	0.319	0.76	0.457	0.85	0.410
	objects > homogenous	1.13	0.276	0.74	0.472	1.06	0.307	1.02	0.325
visible	faces > homogenous	7.74	< 0.001 *	6.06	< 0.001 *	1.91	0.075	3.14	0.007 *
	houses > homogenous	5.83	< 0.001 *	5.51	< 0.001 *	2.63	0.019 *	3.17	0.006 *
	objects > homogenous	4.23	0.001 *	4.47	< 0.001 *	3.05	0.008 *	3.26	0.005 *

Discussion

As in previous studies (Kouider, et al., 2008; Marois, et al., 2004; Moutoussis & Zeki, 2002), this study shows that invisible stimuli can cause stimulus specific activity in object selective parts of the human brain. Moreover, we are able to produce this effect even when the stimuli are task-irrelevant and non-attended, thus suggesting that such categorizing can be established automatically. Additionally, we show that under these conditions there is no significant difference in the level of stimulus specific activity between visible and invisible stimuli, making accounts relying on activity thresholds for visual awareness (e.g. Moutoussis & Zeki, 2002) somewhat unlikely. Apparently, as noted by these and other authors (Beck, Rees, Frith, & Lavie, 2001; Dehaene, et al., 2001), stimulus specific activation (or stimulus categorization by the brain) is insufficient to generate visual awareness. If stimulus specific activity is insufficient to generate visual awareness, what is?

We identified robust non-stimulus specific differences between visible and invisible stimuli in areas of the visual cortex that have previously been implicated in object recognition (LOC, also see

Grill-Spector, et al., 2001) and low level vision (V1 and V2). In these areas, visible stimuli containing a figure-ground relationship generated much more activity than homogenous screens, whereas invisible stimuli containing such a relationship showed no increased levels of activation in this contrast, despite the fact that both visible and invisible stimuli were able to generate stimulus specific activity in higher and/or more specialized areas.

It is especially surprising that this pattern of results was also found in the earliest visual area V1, an area with small receptive fields that has traditionally been found to respond only to simple stimulus characteristics such as line orientation (Hubel & Wiesel, 1962, 1968). Estimates of V1 receptive field size in macaques are ~0.25° in the fovea and less than 1° even at sizeable eccentricities (A. T. Smith, et al., 2001). Estimates of early visual receptive field size in humans using subdural electrodes also point to (much) smaller than 1° sizes (Yoshor, et al., 2007). As Gabor elements in this study subtended a visual angle comparable to the receptive field size of V1 neurons (0.58°) and conditions were balanced out at this level (both monocularly and binocularly, and both within the visible conditions and within the invisible conditions) one would not expect selective differences between conditions to arise here, as was confirmed for invisible stimuli but not for visible stimuli. What then brings about these differences between conditions for visible stimuli in V1? It seems unlikely that these are caused by long-range horizontal intracortical connections. On a monocular level, both visible and invisible stimuli contain figure-ground relationships. If long range horizontal connections would be responsible for the figure-ground signals, it is unclear why such connections would selectively come into play for binocularly driven V1 neurons but not for monocularly driven V1 neurons, each of which comprise approximately 50% of the neurons in V1 (Trotter, 1995).

Alternatively, it may be that recurrent interactions with areas higher up the cortex modulate V1 activity for visible, but not for invisible stimuli. Over the past 15 years, percept-dependent modulation of V1 activity has been repeatedly associated with the influence of feedback connections which have been suggested to play a role in figure-ground segregation (Lamme, 1995), perceptual grouping (Roelfsema, 2006) and visual awareness (Haynes, et al., 2005;

Lamme, 2006). In this interpretation, objects and scenes are categorized quickly and automatically in the feedforward sweep (e.g. Thorpe, et al., 1996; VanRullen & Koch, 2003), whereas conscious representations in which objects acquire their phenomenal figure-ground properties, require recurrent interactions with lower visual areas. Thus a face may be categorized as such in the FFA in the feedforward sweep, but to consciously see the face requires figure-ground segregation and possibly grouping through recurrent interactions.

Perceptual hypothesis testing (Enns & Di Lollo, 2000) may be an integral part of this process, allowing low and high level neurons to be locked into certain perceptual interpretations depending on how well they match. As the match fails, recurrent interactions may fall apart, resulting in a lack of visibility for (part) of the visual input or a transition from one interpretation to another, as may also be the case in pattern masking (Fahrenfort, et al., 2007; Lamme, et al., 2002), flash suppression (Wilke, Logothetis, & Leopold, 2003) and binocular rivalry (Lee, Blake, & Heeger, 2005). In the present study, this would explain why modulatory activity in early and midlevel visual areas is seen for visible, but not for invisible stimuli, as the monocular higher level interpretation does not match the binocular input for invisible stimuli, whereas it does for visible stimuli. For such an explanation to hold, one needs to assume that binocularly plausible interpretations are favored over monocular interpretations, evidence for which exists in the literature (I. Kovacs, Papathomas, Yang, & Feher, 1996).

Our results also have implications for theories of object categorization at large. Traditionally, object categorization is thought to follow figure-ground segregation (Rubin, 1958). More recently, categorization has been thought to influence figure-ground segregation (e.g. Peterson & Gibson, 1993; 1994) or to be based on the same process as figure-ground segregation (Grill-Spector & Kanwisher, 2005; but see Mack, Gauthier, Sadr, & Palmeri, 2008). Here we show that stimulus specific activity can occur even without the neural and behavioral signature of figure-ground segregation. On the basis of these data we are inclined to conclude that initial object categorization is the first (automatic) step in visual processing, and as such precedes figure-ground segregation. This would also explain why

one is able to categorize an object as soon as one becomes aware of its presence (Grill-Spector & Kanwisher, 2005), as in the proposed scheme of events the brain has already categorized the object before it is segregated from the rest of the scene.

At first glance, it may seem puzzling how an object can be categorized without it being segregated from the rest of the image prior to categorization. However, recent modeling efforts have shown the plausibility of strictly feedforward object categorization in natural scenes (Serre, et al., 2007). In addition, incremental grouping theory by Roelfsema (2006; 2000) outlines in some detail how image elements can initially be grouped and classified in feedforward base-groupings causing complex tuning properties in high visual areas, whereas incremental grouping that depends on recurrent interactions enhances the responses of neurons coding features that are bound in perception later on. We hypothesize that in this study, both invisible and visible faces and houses are detected and categorized in feedforward base-groupings, whereas only visible faces and houses are incrementally grouped in recurrent interactions, leading to visual awareness and figure-ground segregation.

Conclusion

The results of this study indicate that dichoptically invisible stimulation is sufficient to generate face and house specific activation, even when stimuli are not task relevant and there is no neural signature of figure-ground segregation. A neural signature of visual awareness and figure-ground segregation can only be found when stimuli are dichoptically visible. This suggests that initial categorization of stimuli is automatic and unconscious and can take place in the absence of visual awareness and figure-ground segregation, plausibly as a result of feedforward processing. Furthermore, these results seem to show that when attention and/or behavioral report of a stimulus is not required, the neural signature of phenomenal visual awareness and figure-ground segregation can be identified in early and midlevel visual areas. It seems likely that this signature is caused by recurrent signals between these areas, reaching all the way back to visual area V1. These findings have important implications for theories of visual awareness and object categorization.