



HAL
open science

On second glance: Still no high-level pop-out effect for faces.

Rufin Vanrullen

► **To cite this version:**

Rufin Vanrullen. On second glance: Still no high-level pop-out effect for faces.. Vision Research, 2006, 46 (18), 3017-27; author reply 3028-35. 10.1016/j.visres.2005.07.009 . hal-00111089

HAL Id: hal-00111089

<https://hal.science/hal-00111089>

Submitted on 10 Nov 2006

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.

On second glance: still no high-level pop-out effect for faces.

Rufin VanRullen

CNRS, Centre de Recherche Cerveau et Cognition,
133 Rte de Narbonne 31062 Toulouse Cedex (France).

Abstract: A recent article in this journal (Hershler & Hochstein, 2005) reported, in contradiction to several earlier studies, that photographs of human faces can be searched for efficiently (i.e. “pop out”) among photographs of other objects (as long as these objects are not “too similar” to faces). An apparent search asymmetry between faces and other categories (houses, cars) pointed to the existence of a specialized “face map”. Findings of impaired performance for scrambled images were presented as evidence that this face pop-out is a high-level, “holistic” effect. While the main pop-out effect cannot be disputed, several choices made in that study in terms of experiment design, analysis and interpretation are questionable. After discussing these issues, I report novel experiments which show that (i) the face pop-out effect can be replicated, but under controlled conditions there is no asymmetry between faces and other objects (cars); (ii) inverting pictures and hence disrupting holistic face processing has only a minor effect on search performance; (iii) finally, search becomes inefficient when Fourier amplitude information (which carries global low-level statistical properties of images) is made irrelevant, and only phase information (carrying contour localization) can be used to detect faces. These results imply, contrary to the target article, that the face pop-out effect is mostly based on low-level factors.

1. Introduction

Faces hold a special status for vision scientists: to realize this, one need only count the host of studies debating whether faces hold a special status for our visual systems! While it is undeniable that large numbers of visual neurons respond to faces (Gross, Rocha-Miranda, & Bender, 1972; Perrett, Rolls, & Caan, 1982; Kanwisher, McDermott, & Chun, 1997), more so than for any other category (Chao, Martin, & Haxby, 1999; Haxby, Hoffman, & Gobbini, 2000; Haxby et al., 2001; Kreiman, Koch, & Fried, 2000), we still do not know whether this apparent preference is confounded by exposure or expertise factors (in short, we simply see more faces than other objects; (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Gauthier, Skudlarski, Gore, & Anderson, 2000; Tarr & Gauthier, 2000; Rossion, Kung, & Tarr, 2004; Diamond & Carey, 1986)). Another open question is whether faces are processed faster than other categories of objects, and possibly through different routes (Farah, Wilson, Drain, & Tanaka, 1998; Morris, Ohman, & Dolan, 1999; Tovee, 1998; Pourtois, Grandjean, Sander, & Vuilleumier, 2004; Vuilleumier, Armony, Driver, & Dolan, 2003; Vuilleumier, Richardson, Armony, Driver, & Dolan, 2004). Yet another important question has been with regard to attentional aspects of face processing: do faces pop out? In other words, are faces searched for efficiently (in “parallel”) during a visual search experiment? Over the last two decades, this question was asked and answered several times. After some hesitation –due to confounds in experimental conditions (Hansen & Hansen, 1988; Purcell, Stewart, & Skov, 1996), the settled answer turned out to be a resounding “no” (Brown, Huey, & Findlay, 1997; Nothdurft, 1993; Wolfe, 1998). Now this question has been addressed again in a recent issue of this journal (Hershler & Hochstein, 2005). The authors used natural photographs or line drawings of faces and other objects to demonstrate the existence of a “pop-out” effect for faces (i.e. a search slope on face-present trials of about 6 ms/item). In order to demonstrate that the observed pop-out is a high-level effect, they relied on the following evidence. First, pop-out was observed for faces among other categories such as cars, but not in the reversed search. This so-called search asymmetry (Treisman & Souther, 1985)

is taken as the reflection of the existence of a specific “feature map” for detecting the presence of faces. Second, face search performance was found to decrease strongly when images were scrambled (i.e. divided in a number of squares that were shuffled randomly), and high-level “holistic” processing was assumed to be disrupted. From here it was inferred that face pop-out is a high-level process. This interpretation is disputed below, together with more specific points raised by the authors. I then present an experimental replication of these results, and further experiments showing that the face pop-out effect exists, but that it can be explained in large part as a consequence of low-level differences between faces and other images.

It is important to point out, however, that the present argument does in no way invalidate the “Reverse Hierarchy Theory” advocated by the authors (Hochstein & Ahissar, 2002). One major prediction of that theory, that high levels of representation can be attained efficiently by the visual system, even in the absence of attention, is something that we and other authors have also demonstrated repeatedly, for faces (Reddy, Wilken, & Koch, 2004) as well as other object categories (Fei-Fei, VanRullen, Koch, & Perona, 2005; Li, VanRullen, Koch, & Perona, 2002; Rousselet, Thorpe, & Fabre-Thorpe, 2004). The only point of disagreement here is on the fact that such preattentive performance for high-level categories cannot guarantee pop-out (Reddy, VanRullen, & Koch, 2005; VanRullen, Reddy, & Koch, 2004; VanRullen, Reddy, & Fei-Fei, 2005). According to our views, this is simply due to local competition for neuronal resources (Desimone & Duncan, 1995; Reynolds & Desimone, 1999) within the large receptive fields underlying high-level processing of object categories (e.g. face processing; (Gross et al., 1972; Perrett et al., 1982)).

2. A critical look at Hershler and Hochstein’s arguments.

2.1. Faces pop out, but only among non-face-like objects?

Hershler and Hochstein start by acknowledging the numerous previous reports which failed to observe pop-out for faces. They rationalize these failures by the fact that distractors in those studies had been too similar to faces for the face detection mechanism to discard them. With a similar reasoning however, just about any visual category could be made to pop out, simply by making distractors more and more different from it. A picture of a red car can easily be made to pop out, for example, if one ensures that none of the distractors will be red, or will have a horizontally elongated shape with sharp angles. In such a case, can one claim that cars pop out among other objects, or should one be more cautious and say that red objects pop out among non-red objects, elongated shapes among non-elongated shapes, etc? One influential account of visual search (Duncan & Humphreys, 1989) insists on target-distractor dissimilarity and distractor-distractor similarity as key elements in the generation of pop-out effects. In this theory similarity and dissimilarity are defined in terms of low-level features. The fact that faces pop out only when they are visually dissimilar to other objects, as promoted by Hershler and Hochstein, can be viewed as a mere confirmation of this 15 year-old theory.

2.2. No speed-accuracy trade-off?

In their analysis, Hershler and Hochstein measure the reaction time slope for target-present trials only. This is debatable, but not exactly unconventional (Treisman & Souther, 1985). The slopes are found to be less than 10 ms per element for faces, which is indicative of parallel (or “efficient”) search (Wolfe, 1998), since the additional processing time taken for each stimulus in the display can be considered fairly small. One needs to verify, however, that this efficiency is not due to a speed-accuracy trade-off strategy, whereby the subjects would react fast but inaccurately on trials with large set sizes. For most of their results, the authors find no effect of set size on percent correct or d' . The problem here is that percent correct or d' are calculated over both target-present and target-absent trials, whereas reaction times were estimated from target-present trials only. The fact that there is no speed-accuracy trade-off might simply stem from this

unfair comparison. Obviously the fair thing to do would have been to calculate percent correct only for target-present trials, and then test for an effect of set size. This might seem to be a minor concern, but it can in fact have a serious impact on the conclusions, as developed below.

2.3. Larger set sizes mean more accurate slopes?

Hershler and Hochstein insist that their using larger set sizes (here up to 64 pictures) than most visual search experiments (usually limited to 10 or 20 pictures) is one reason for their success in showing a pop-out effect for faces. More images to analyze would yield better estimates of visual search slopes. This argument is, unfortunately, invalid: as pointed out for example by Chun and Wolfe (Chun & Wolfe, 1996), search will often terminate before all items are examined, resulting in artificially flat slopes, and this problem will be more apparent with larger set sizes. To give an extreme example, imagine a subject presented with 10,000 pictures on a large wall and asked to look for a face. When the subject has not found a target after about a minute of examination, he/she is likely to give up the search and answer "no target" (especially when they know that a few hundred such trials are to follow). Would the observed search slope (about 6 ms/item) really imply that search was parallel? Of course, in this extreme example performance on target-present trials would certainly drop dramatically at larger set sizes, giving away the subject's stratagem. But given what was said in the previous point (2.2), this speed-accuracy trade-off might go unnoticed if target-absent trials (for which performance will tend to become near-perfect at large set sizes, when the subject grows increasingly likely to answer "no target") are taken into account.

2.4. Search asymmetry?

The search asymmetry between faces and other object categories (cars, houses), which constitutes an essential portion of Hershler and Hochstein's demonstration, was obtained under conditions that did not favor optimal comparison between categories. Within a face search block, distractors could be either cars or houses (this was pseudo-randomly decided on each trial); within a car search block, distractors could be faces or houses, etc. One need only assume that cars and houses share one or more simple low-level features (sharp angles, horizontal and vertical lines) not shared by face drawings to explain the observed "asymmetry". On face search blocks, this difference in simple features could have supported (a "low-level") pop-out on every trial. But on car search blocks (or house search blocks), the same features would only turn out to be helpful on half the trials (those having faces as distractors). In order to preserve a decent accuracy level over the entire block, subjects might have chosen to switch from an unfruitful "parallel" to a more conservative "serial" strategy. In fact, it does appear that in car-among-face search trials performance was much better than for car-among-house trials (13 ms/item in the former, 21 in the latter). These inefficient car-among-house trials might have had damaging consequences on the search strategy adopted by subjects over the entire block. The same rationalization was in fact used by Hershler and Hochstein to explain the poor performance (slope of 17 ms/item) on non-scrambled face search trials in their experiment 4, and there is no reason why it could not apply here as well. To conclude, the existence of a search asymmetry should be tested under fair experimental conditions, in which blocks of, say, car-among-face search are compared to blocks of face-among-car search. This is precisely what was tested in my experiment 2 (reported below) and the results did not concur with those of Hershler and Hochstein.

2.5. Scrambling disrupts parallel search (evidence for holistic processing)?

The points listed above challenge the conceptual design, experimental and analysis methods used to conclude that a pop-out effect exists specifically for face images. Even assuming that this first conclusion could be validated, however, there remains a second debatable conclusion on Hershler and Hochstein's agenda: showing that this pop-out reflects a true, high-level effect. For this, they use a scrambling procedure, dividing each picture (target and distractor alike) into 4, 9 or 16 squares, and permuting these squares randomly. The resulting stimuli do not resemble the

original: a scrambled face might contain preserved face parts (a nose, an eye) but could not support holistic recognition of the face. When the intensity of the scrambling increased (from 4 to 16 squares), search performance decreased dramatically (with slopes as high as 51ms/item). This was interpreted as evidence that the pop-out effect is a high-level, holistic one: when holistic processing is disrupted by the scrambling procedure, pop-out disappears.

Unfortunately, the chosen procedure does not only disrupt holistic processing, but also makes the face and distractor stimuli more similar in their low-level properties, for example in terms of their Fourier amplitude spectrum (by introducing large amounts of spurious energy in the horizontal and vertical orientations). Here again, the well-known target-distractor similarity factor proposed by Duncan and Humphreys (Duncan & Humphreys, 1989) might suffice to explain the loss of performance.

In addition, there are other reasons why this scrambling manipulation should be interpreted cautiously. First, it might be that this manipulation increases effective set-size, such that when subjects are searching through an array of 64 images scrambled into 16 parts each, they might effectively be searching through 1024 "items". This multiplication of elements could easily explain the observed 10-fold increase in the slope for scrambled pictures. Second, a recognition based on face parts (as opposed to true holistic recognition) does not necessarily rely on obvious facial features such as "an eye" or "a mouth" (as proposed by Hershler and Hochstein), but could also be based on larger face parts including pairs or triplets of such features (Ullman, Vidal-Naquet, & Sali, 2002). That an isolated mouth or eye is not sufficient for parallel recognition does not, therefore, directly imply a "holistic" or "high-level" effect in parallel face detection (in other words, intermediate situations between full-blown "holistic" recognition and recognition by "simple parts" need to be considered as well).

In one of the following experiments, I use a simpler, and more classical way of disrupting holistic processing: the face inversion effect (Yin, 1969; Valentine, 1988; Farah, Tanaka, & Drain, 1995). When an inverted face is searched for among inverted pictures of other objects, search performance remains efficient even though holistic processing is disrupted. This is again in favor of a low-level account for this pop-out effect.

3. Replication and further experiments

METHODS.

10 subjects (4 females) aged between 23 and 34 years old, with normal or corrected-to-normal vision, participated in these visual search experiments (except for Experiment 3, where only 6 of the previous subjects were tested). Experiments 1, 2 and 4 were performed within the same session, in random order (counterbalanced across subjects). Experiment 3 was performed in a later separate session. For each subject and condition, 198 search trials were collected. Three set sizes were used (16, 32 or 64 items), so the results would be comparable to those obtained by Hershler and Hochstein. At each set size, the pictures (occupying approximately 2x2 degrees of visual angle) were shown on a black screen, at a random position on a virtual 8x8 grid (except for set size 64 where all positions were occupied). Half the trials contained a target picture, and the subjects were asked to press one of two mouse buttons, as fast as possible, to indicate whether the target was present or not. They were instructed to maintain error rates within 10% in all conditions. Auditory feedback was given on erroneous trials. As in Hershler and Hochstein, only the search slopes (average reaction time variation per display element) for correct target-present trials were analyzed and discussed here, as they are assumed to better reveal the parallel/serial nature of the search task (Treisman & Souther, 1985). However, we also report the target-absent slopes and the accuracy data (d') for completeness. The picture database was obtained mainly from the Corel CD-ROM libraries and a few additional digitized pictures. It consisted of at least 109 pictures of each of the following categories: faces, cars, other means of transport, animals, other natural scenes. In experiment 1, subjects were asked to search for a face among instances of all other categories ("non-faces"), or (on different blocks) a car among instances of all other categories ("non-cars"). In experiment 2, subjects searched for a face among cars or (on different blocks) a car among faces. In experiment 3, subjects searched for an inverted (180° rotation) face among inverted non-faces. In experiment 4 subjects were asked to look for a face among non-face

objects, and the pictures were manipulated as described below, so that Fourier amplitude information could be used (face/face hybrid condition) or not (face/car hybrid condition) to perform the search task.

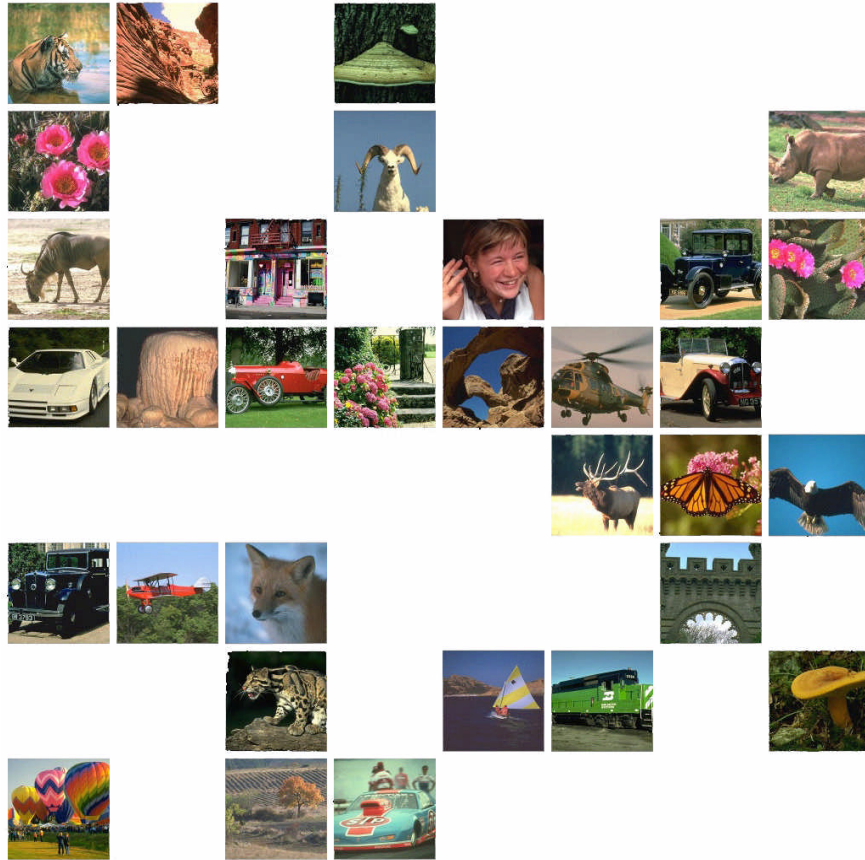


Figure 1. Example of a face-among-non-face search trial in Experiment 1. In this example the set size is 32 and the target is present. In the actual experiment images were presented in color on a black background.

EXPERIMENT 1: Replication.

A set of natural photographs was tested under conditions similar to Hershler and Hochstein's experiment (Figure 1). When the target was a face image among various pictures including animals, cars, other means of transport, landscapes etc, search performance was parallel (Figure 2; target-present slope 3 ms/item). When the target was a car under similar conditions (with distractors including faces, animals, other means of transport etc.), search slopes were very high in comparison (Figure 2; target-present slopes 22 ms/item). A 2-way ANOVA (set size x task) confirmed a main effect of set size on reaction times ($F(2,54) = 9.4, p < 0.001$), a main effect of task ($F(1,54) = 86.3, p < 0.001$) and a significant interaction ($F(2,54) = 5.3, p < 0.01$). In addition, a comparison of individual search slopes revealed that they were significantly higher for the car than for the face search (t-test, $t(18) = 3.6, p < 0.01$). This essentially replicates Hershler and Hochstein's main conclusions: faces appear to pop out among non-faces, and this pop-out effect appears to be specific to this category of objects.

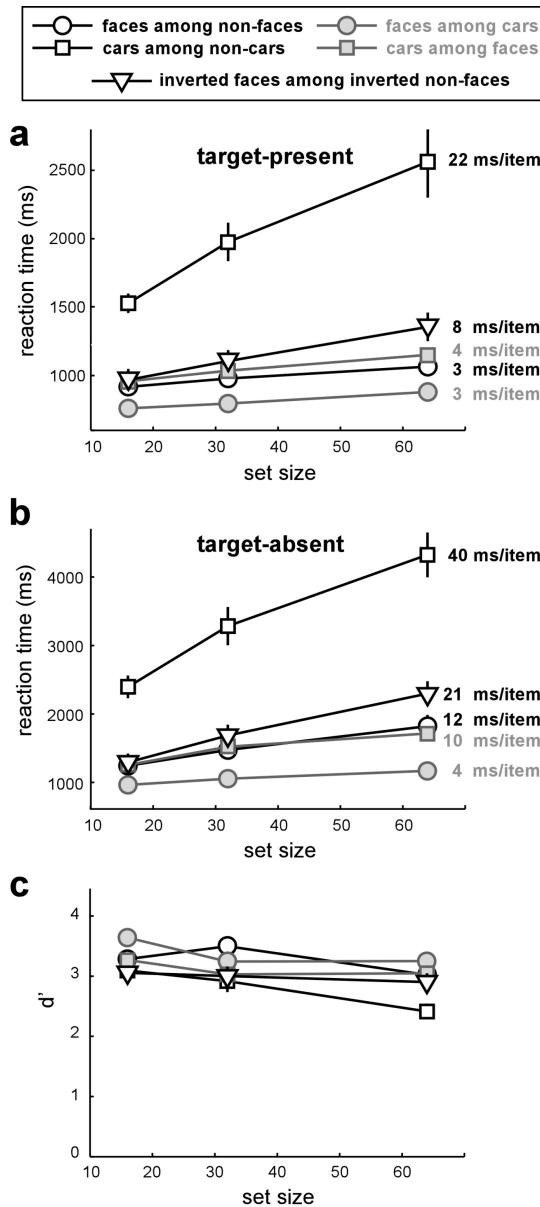


Figure 2. a. Target-present slopes for Experiments 1, 2 and 3. The main pop-out effect from Hershler and Hochstein’s study was replicated here (Expt 1): face-among-non-faces search was efficient, with slopes of 3 ms/item, whereas car-among-non-cars search slopes were much higher (22 ms/item). This advantage for faces vanished when it was tested with more controlled distractors (Expt 2): face-among-cars search slopes were 3 ms/item, but car-among-faces search slopes were only 4 ms/item. Finally, when using inverted images (Expt 3) and thus disrupting holistic processing, search remained fairly efficient (8 ms/item). Target-absent slopes (**b.**) and d' accuracy measures (**c.**) are also shown for completeness, and display a pattern of results compatible with these conclusions. Note the difference in the y-axis scale between panels **a** and **b**.

EXPERIMENT 2: search asymmetry?

As pointed out above, search asymmetry should be tested under controlled conditions, where entire blocks of face-among-cars search are compared with blocks of car-among-faces search. When subjects searched for cars among (only) face distractors, search was just as efficient (Figure 2; target-present slope 4 ms/item) as when the search was for faces among (only) cars (Figure 2; target-present slope 3 ms/item). A 2-way ANOVA revealed significant main effects of set size ($F(2,54)=9.7$, $p<0.001$), task ($F(1,54)=65.0$, $p<.001$), but no significant interaction ($F(2,54)=0.5$, $p>0.05$). Search slopes were not significantly different in these two conditions (t -test, $t(18) = 0.6$, $p>0.05$). This implies that there is no real asymmetry between the two categories, and suggests that the pop-out effect might simply be explained by some low-level difference between face images and other images. This is further tested in the following experiments.

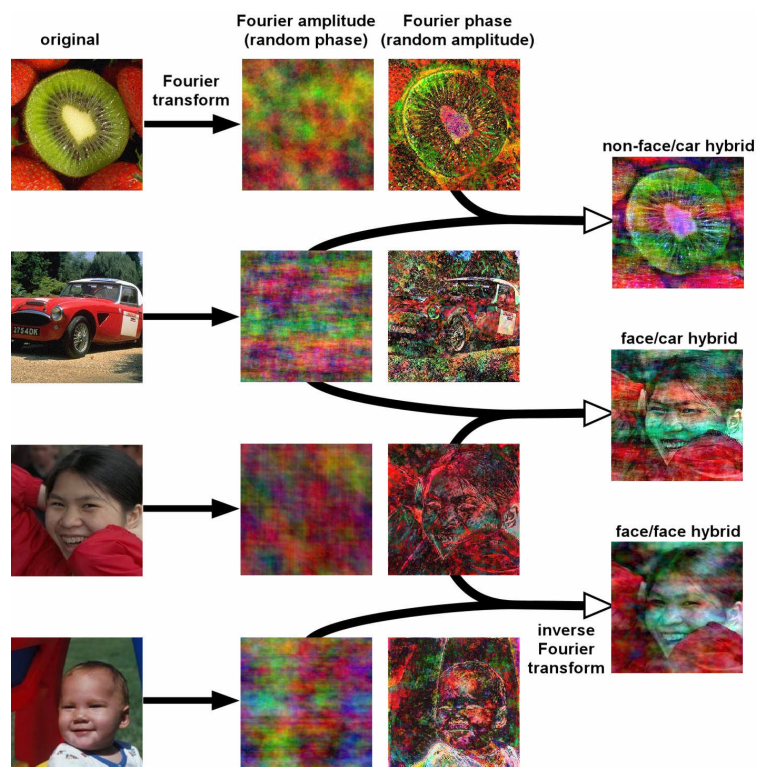


Figure 3. Examples of Fourier-based image manipulations and hybrid images. The Fourier transform separates phase from amplitude information within natural images. Fourier amplitude information reflects several low-level aspects of the original picture, such as the global distribution of orientations and spatial frequencies. Fourier phase information, on the other hand, captures important high-level properties of the original scene, such as contour localization information. One can create various types of hybrid images by recombining (using the inverse Fourier transform) phase and amplitude information from different images. Using these hybrid images, one can study search performance when (low-level) amplitude information can or cannot be utilized to perform the search task (Experiment 4).

EXPERIMENT 3: face inversion and holistic processing.

Face inversion is known to impair holistic face processing not only in tasks involving face recognition, but also during face detection (Purcell & Stewart, 1988; Valentine & Bruce, 1986; Lewis & Edmonds, 2003; George, Jemel, Fiori, & Renault, 1997; Rodriguez et al., 1999). When asked to detect an inverted face among inverted non-face objects, the subjects' search performance was slightly impaired compared to the corresponding condition with upright stimuli (Experiment 1), but remained efficient in comparison with the other conditions (Figure 2; target-present slope 8ms/item). A 2-way ANOVA comparing reaction times for upright and inverted face search indicated a significant main effect of set size ($F(2,30)=5.3$, $p=0.01$), but no effect of task ($F(1,30)=2.73$, $p>0.05$), and no significant interaction ($F(2,30)=0.7$, $p>0.05$). Search slopes were not significantly different in these two conditions (t-test, $t(10) = 1.6$, $p>0.05$). This is important because the face inversion effect implies that "holistic" processing of the face was disrupted under these conditions. Therefore, and contrary to Hershler and Hochstein's conclusions, the face pop-out effect does not appear to reflect high-level or "holistic" parallel processing of faces.



Figure 4. Example of a face/car hybrid among non-face/car hybrids search trial in Experiment 4. Because all pictures in the search array have been given the Fourier amplitude spectrum of a car image (different for each picture), search performance must rely on Fourier phase information (e.g. contour localization). In this example the set size is 32 and the target is present (bottom left). In the actual experiment images were presented in color on a black background.



Figure 5. Example of a face/face hybrid among non-face/car hybrids search trial in Experiment 4. Distractors are similar to those described in Figure 4, but here the face target (left) is the only image with the amplitude spectrum of a face. Amplitude information can thus be used to perform this search. In the actual experiment images were presented in color on a black background.

EXPERIMENT 4: Fourier amplitude vs. phase.

In order to understand whether low-level factors can be responsible for the previous pop-out effect (Experiment 1), datasets of hybrid images were created in which the Fourier phase information of one picture was combined with the Fourier amplitude information of another (Figure 3). The resulting hybrid image is always categorized according to the phase information (i.e. it resembles a face when the phase information of a face is used, but not when the phase information of another object is combined with the amplitude of a face). While there is no a priori theoretical reason to associate phase with high-level, and amplitude with low-level information, a brief look at figure 3 suffices to appreciate this empirical correspondence. In addition, phase-scrambled images (with intact amplitude spectrum) are often used as a baseline condition to isolate high-level visual areas in fMRI experiments, because low-level striate and extra-striate areas continue to respond to these images (Rainer, Augath, Trinath, & Logothetis, 2001; Olman, Ugurbil, Schrater, & Kersten, 2004; Grill-Spector et al., 1998) while high-level activity (e.g. in temporal cortex) is disrupted (Malach et al., 1995; Grill-Spector et al., 1998; Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003; Sawamura, Georgieva, Vogels, Vanduffel, & Orban,

2005). This supports the idea that the phase spectrum of natural images can be related to high-level visual information, whereas the amplitude spectrum carries lower-level information.

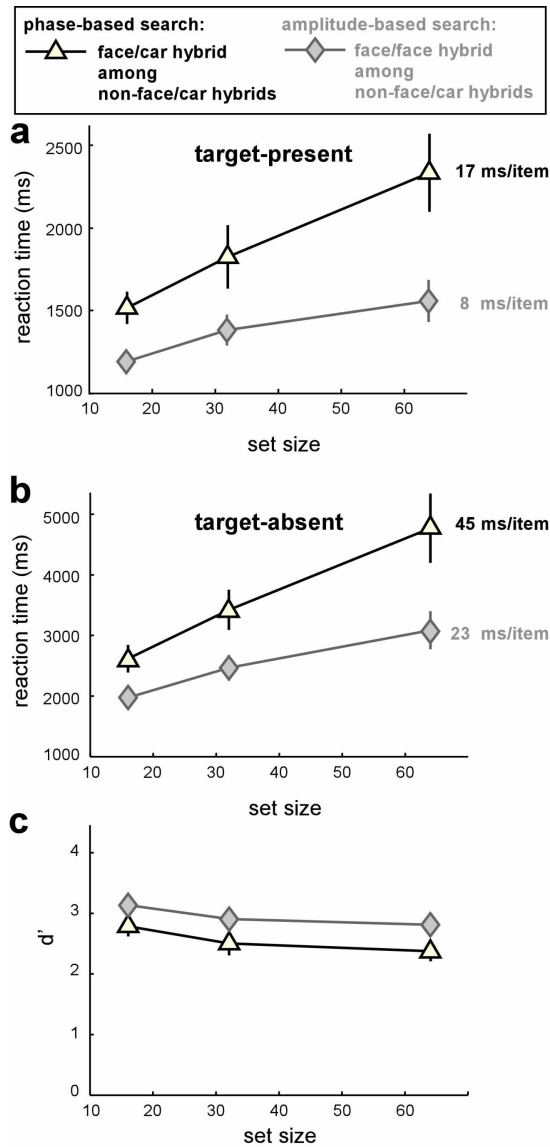


Figure 6. a. Target-present slopes for Experiment 4. The slopes, obtained with hybrid images as described in Figure 3, reflect what happens when low-level amplitude information can be used (8 ms/item) or not (17 ms/item) to perform the search. Target-absent slopes (**b.**) and d' accuracy measures (**c.**) are also shown for completeness. Note the difference in the y-axis scale between panels **a** and **b**.

Hybrid images (Figure 3) were used in a search experiment to determine the contribution of low-level factors comprised in the Fourier amplitude spectrum. In one condition, subjects searched for a face/car hybrid among non-face/car hybrids. Every picture (target and distractor alike) had the amplitude spectrum of a car (a different instance of car was used for each of the hybrid images, so that the resulting images were not too uniform), and so amplitude information could not be used to perform the search (Figure 4). In that case, search became quite inefficient

(Figure 6; target-present slope 17 ms/item). Note that this condition is identical to the parallel search obtained in Experiment 1 in all respects, but for the availability of amplitude-based information. Thus it appears that high-level phase information alone cannot underlie the parallel search for faces. Of course, it is still possible that this performance impairment might be due to the artificial and “fuzzy” appearance of the hybrid images. But when subjects searched for a face/face hybrid among non-face/car hybrids (i.e. the same condition as before, except that amplitude information was now available to detect the target; Figure 5), search became much more efficient (Figure 6; target-present slope 8 ms/item). A 2-way ANOVA comparing reaction times for these two hybrid search tasks indicated significant main effects of set size ($F(2,54)=7.4$, $p<0.01$) and task ($F(1,54)=16.5$, $p<0.001$), but no significant interaction ($F(2,54)=1.1$, $p>0.05$). However, individual search slopes were significantly different between these two conditions (t -test, $t(18) = 2.3$, $p<0.05$). Thus it is not the image manipulation itself that is responsible for the task difficulty. Rather, the critical difference between the former inefficient and the latter efficient face search conditions lies in the amplitude information (e.g. compare the 2nd and 4th images in the second column of Figure 3).

Although it is unclear exactly what aspects of the amplitude spectrum underlie this parallel search, and what mechanisms might implement this in the brain (in particular, none of the present results should be taken to imply that the visual system explicitly extracts the Fourier amplitude or phase spectra from natural images), these results suggest that Hershler and Hochstein’s claim that this pop-out is a “high-level” effect might be somewhat exaggerated.

4. Discussion

As in their discussion of previous experimental failures to show a pop-out effect for faces, Hershler and Hochstein might argue in response to the present demonstration that the image manipulation (Experiment 4) simply made the face target and non-face distractors too similar to be efficiently discriminated. I will not object here, since this was exactly the purpose of the manipulation. But importantly, this manipulation only affected low-level information contained in the Fourier amplitude spectrum. Targets and distractors were made more similar in terms of their low-level properties, and search became inefficient, as could be predicted from the prevalent “similarity-dissimilarity” theory of Duncan and Humphreys (1989). High-level, phase-related contour information did not suffice to salvage the pop-out effect. Hence, Hershler and Hochstein’s conclusion that face pop-out is a high-level effect must be rejected.

Again, as pointed out in the introduction, this should not be taken to invalidate the Reverse Hierarchy Theory (RHT) of visual processing defended by Hochstein and Ahissar (2002). There are other large bodies of evidence demonstrating that high-level categorization of objects and scenes can proceed in the absence of attention (Li et al., 2002; Rousselet, Fabre-Thorpe, & Thorpe, 2002; Reddy et al., 2004; Fei-Fei et al., 2005), which is precisely the prediction of RHT that was being tested by Hershler and Hochstein. The low-level confound identified here is unlikely to apply to these previous results: for example, Reddy and colleagues (Reddy et al., 2004) reported that carefully controlled male and female faces could be distinguished preattentively, and Fourier amplitude information certainly does not differ significantly between these two categories of images; similarly, cars can be discriminated from other objects when attention is occupied elsewhere (Li et al., 2002), even though low-level differences between these categories of images were not sufficient in the present Experiment 1 to support a pop-out effect. In other words, it seems that, although high-level natural object discriminations do not require significant attention, the same high-level discriminations are not performed in parallel in visual search experiments (VanRullen et al., 2004; VanRullen et al., 2005; Reddy et al., 2005; Rousselet, Thorpe, & Fabre-Thorpe, 2004), at least when low-level factors are controlled for. Why might this be?

We have argued previously (Reddy et al., 2005; VanRullen et al., 2004; VanRullen et al., 2005) that visual search performance for high-level properties will be necessarily limited by local competition factors, due to the large size of neuronal receptive fields involved in high-level object (and face) representations (see (Rousselet et al., 2004) for an up-to-date review of receptive fields sizes along the ventral pathway). Such competition is known to weaken neuronal representations (Desimone & Duncan, 1995; Reynolds & Desimone, 1999), and attention becomes necessary to restore neuronal activity. The search task becomes serial, even when each object in isolation would not have required attention. In practice, this simple and straightforward logic implies that there can never be a truly “high-level” parallel search, and suggests that any report of a “high-level pop-out” effect such as the one by Hershler and Hochstein (2005) must conceal one or more low-level confounds, as demonstrated here.

5. Acknowledgements

This work was supported by the CNRS. I wish to thank Holle Kirchner and Guillaume Rousselet for helpful discussions, and Lavanya Reddy, Leila Reddy, Jeremy Wolfe and Adam Reeves (Editor) for valuable comments on the manuscript.

6. References

- Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: do faces pop out? *Perception*, *26*(12), 1555-1570.
- Chao, L. L., Martin, A., & Haxby, J. V. (1999). Are face-responsive regions selective only for faces? *Neuroreport*, *10*(14), 2945-2950.
- Chun, M. M., & Wolfe, J. M. (1996). Just say no: how are visual searches terminated when there is no target present? *Cognit Psychol*, *30*(1), 39-78.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193-222.
- Diamond, R., & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *J Exp Psychol Gen*, *115*(2), 107-117.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychol Rev*, *96*(3), 433-458.
- Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *J Exp Psychol Hum Percept Perform*, *21*(3), 628-634.
- Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychol Rev*, *105*(3), 482-498.
- Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition*, *in press*.
- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci*, *3*(2), 191-197.
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci*, *2*(6), 568-573.
- George, N., Jemel, B., Fiori, N., & Renault, B. (1997). Face and shape repetition effects in humans: a spatio-temporal ERP study. *Neuroreport*, *8*(6), 1417-1423.
- Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzhak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp*, *6*(4), 316-328.

- Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *J Neurophysiol*, 35(1), 96-111.
- Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: an anger superiority effect. *J Pers Soc Psychol*, 54(6), 917-924.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539), 2425-2430.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn Sci*, 4(6), 223-233.
- Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Res*, 45(13), 1707-1724.
- Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5), 791-804.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*, 17(11), 4302-4311.
- Kreiman, G., Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat Neurosci*, 3(9), 946-953.
- Lewis, M. B., & Edmonds, A. J. (2003). Face detection: mapping human performance. *Perception*, 32(8), 903-920.
- Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci U S A*, 99(14), 9596-9601.
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A*, 92(18), 8135-8139.
- Morris, J. S., Ohman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc Natl Acad Sci U S A*, 96(4), 1680-1685.
- Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. *Perception*, 22(11), 1287-1298.
- Olman, C. A., Ugurbil, K., Schrater, P., & Kersten, D. (2004). BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision Res*, 44(7), 669-683.
- Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47, 329-342.
- Pourtois, G., Grandjean, D., Sander, D., & Vuilleumier, P. (2004). Electrophysiological correlates of rapid spatial orienting towards fearful faces. *Cereb Cortex*, 14(6), 619-633.
- Purcell, D. G., & Stewart, A. L. (1988). The face-detection effect: configuration enhances detection. *Percept Psychophys*, 43(4), 355-366.
- Purcell, D. G., Stewart, A. L., & Skov, R. B. (1996). It takes a confounded face to pop out of a crowd. *Perception*, 25(9), 1091-1108.
- Rainer, G., Augath, M., Trinath, T., & Logothetis, N. K. (2001). Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Curr Biol*, 11(11), 846-854.
- Reddy, L., VanRullen, R., & Koch, C. (2005). Inter-stimulus distance effects in visual search. (*submitted*).
- Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *Journal of Vision*, 4(2), 106-117.
- Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, 24(1), 19-29, 111-125.
- Rodriguez, E., George, N., Lachaux, J. P., Martinerie, J., Renault, B., & Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature*, 397(6718), 430-433.

- Rossion, B., Kung, C. C., & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proc Natl Acad Sci U S A*, *101*(40), 14521-14526.
- Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nat Neurosci*, *5*(7), 629-630.
- Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends Cogn Sci*, *8*(8), 363-370.
- Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: the limits of parallelism. *Vision Res*, *44*(9), 877-894.
- Sawamura, H., Georgieva, S., Vogels, R., Vanduffel, W., & Orban, G. A. (2005). Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. *J Neurosci*, *25*(17), 4294-4306.
- Tarr, M. J., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci*, *3*(8), 764-769.
- Tovee, M. J. (1998). Is face processing special? *Neuron*, *21*(6), 1239-1242.
- Treisman, A., & Souther, J. (1985). Search asymmetry: a diagnostic for preattentive processing of separable features. *J Exp Psychol Gen*, *114*(3), 285-310.
- Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nat Neurosci*, *6*(9), 989-995.
- Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nat Neurosci*, *5*(7), 682-687.
- Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *Br J Psychol*, *79* (Pt 4), 471-491.
- Valentine, T., & Bruce, V. (1986). The effects of distinctiveness in recognising and classifying faces. *Perception*, *15*(5), 525-535.
- VanRullen, R., Reddy, L., & Fei-Fei, L. (2005). Binding is a local problem for natural scenes and objects. *Vision Res*, (in press).
- VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual-tasks reveal two distinct attentional resources. *J Cog Neuroscience*, *16*(1), 4-14.
- Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nat Neurosci*, *6*(6), 624-631.
- Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nat Neurosci*, *7*(11), 1271-1278.
- Wolfe, J. M. (1998). Visual Search. In H. Pashler (Ed.), *Attention* (pp. 13-73). London, UK: University College London Press.
- Yin, R. K. (1969). Looking at upside-down faces. *J Exp Psychol*, *81*(1), 141-145.