Perception, 2010, volume 39, pages 1125 – 1141

doi:10.1068/p6659

Does Thompson's Thatcher Effect reflect a face-specific mechanism?

Yetta K Wong
Department of Psychology, Vanderbilt University, Nashville, TN 37203, USA;
e-mail: yetta.wong@vanderbilt.edu
Elyssa Twedt
Department of Psychology, University of Virginia, Charlottesville, VA, USA
David Sheinberg
Department of Neuroscience, Brown University, Providence, RI, USA
Isabel Gauthier
Department of Neuroscience, Vanderbilt University, Nashville, TN 37203, USA
Received 1 February 2010, in revised form 7 July 2010

Abstract. The Thatcher Illusion or Thatcher Effect (TE—Thompson 1980, *Perception* **9** 483 – 484) reflects the difficulty in perceiving the local inversion of parts when the whole object, generally a face, is globally inverted. We tested the generality of the TE with a range of faces and non-face objects, and observed the TE with many non-face categories including cars, buildings, bikes, and letter strings. In terms of magnitude, the face TE is not exceptionally large compared to other object categories, and the magnitude of the TE can be predicted by performance on this task for upright stimuli, regardless of whether the object is a face or not. We did not observe evidence for a unique mechanism contributing to the TE for faces. We discuss factors that influence the magnitude of the TE, some common across domains and others more specific to a particular category.

1 Introduction

Thompson (1980) reported a striking perceptual phenomenon using the face of Margaret Thatcher. Her face was rendered grotesque by locally inverting the eyes and mouth, but, surprisingly, this transformed face did not look particularly unusual when viewed upside-down (figure 1). This phenomenon was subsequently termed the Thatcher illusion or 'Thatcher Effect' (TE).

In the literature and in the popular media, the TE has often been regarded as a face-specific phenomenon. For instance, this phenomenon is explained by the configural or holistic processing of faces, a strategy that is thought to be unique to faces (Bartlett and Searcy 1993; Bertin and Bhatt 2004; Boutsen and Humphreys 2003; Boutsen et al 2006; Donnelly and Hadwin 2003; Edmonds and Lewis 2007; Lewis 2001; Lewis and Johnston 1997; Rouse et al 2004). Under this explanation, configural information is used during the perception of upright but not inverted faces. Therefore, the local inversion of parts (a configural change) for upright faces is very prominent, but that for inverted faces is much less noticeable (even though the features themselves can still be recognized). In addition, the grotesque expression is considered the most salient quality of the TE such that studies have emphasized the importance of the grotesque expression in thatcherized faces (Murray et al 2000; Parks et al 1985; Rock 1988; Valentine and Bruce 1985). The emotional content carried by the 'grotesqueness' dimension is difficult to apply to non-face objects, again offering indirect evidence that the TE is face-specific.

However, the face-specific view of the TE is not well supported with empirical evidence. First of all, the TE is not unique for faces, which suggests that the TE is more than merely a misperception of facial expression. For example, a TE with words

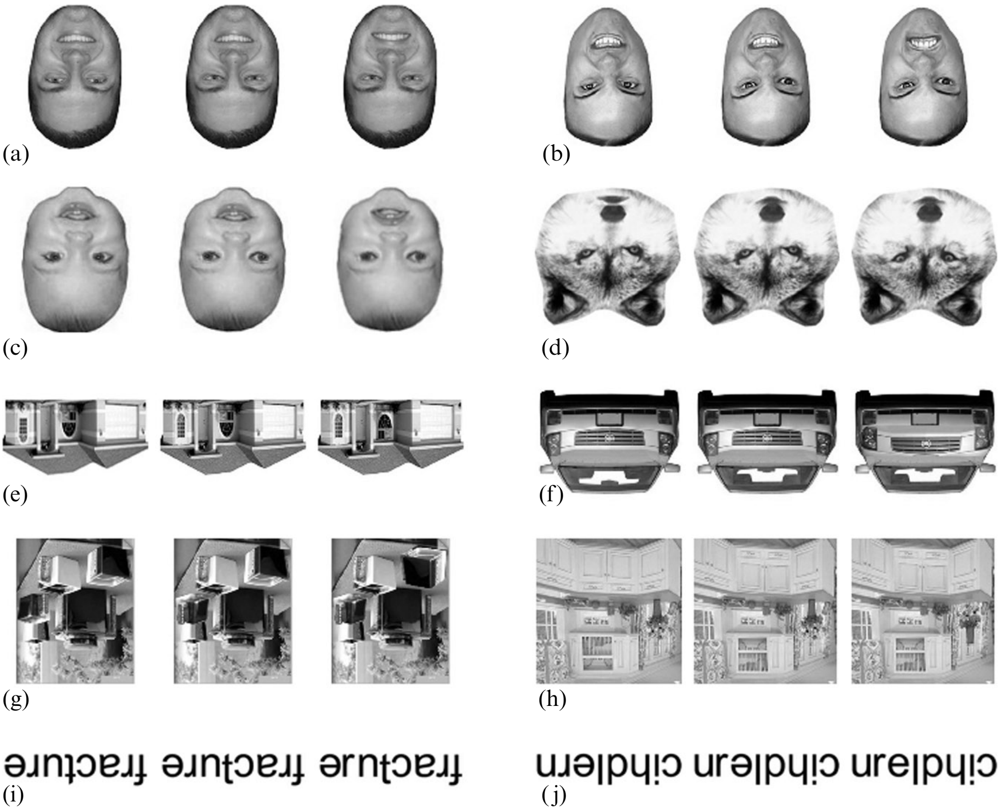

Figure 1. Examples of thatcherization for (a) human faces, (b) grimacing faces, (c) baby faces, (d) animal faces, (e) buildings, (f) cars, (g) close-up scenes, (h) scenes, (i) words, and (j) non-words. In all image groups, the left-hand image is normal, the middle image has been thatcherized by locally inverting one part (thatcherized 1), and the right-hand image has been thatcherized by locally inverting two parts (thatcherized 2). (a) Represents the classical Thatcher illusion for human faces where the eyes and mouth have been locally inverted.

has been illustrated in simple demonstrations (Parks 1983; Rock 1988), and a TE with houses has been previously reported (Boutsen et al 2006; Rouse et al 2004). In addition, there is no empirical evidence supporting that the TE for faces is of an exceptionally large magnitude compared to other object categories. It is, however, theoretically possible that a modest TE can be obtained for all non-face objects owing to domain-general processes (that are sensitive to object orientation), while a larger TE for faces can be observed because of the unique contribution of configural processing. In this study, we assessed the face specificity of the TE by (i) testing, in the same task, whether the TE can be obtained with a range of object categories, including both faces and non-face objects; and (ii) comparing the magnitude of the TE between faces and several non-face object categories. If the TE is specific for faces, we should observe a TE that is unique for faces, and/or a TE that is exceptionally large for faces compared to other object categories.

Studying the TE across a range of objects requires a task different from those used in previous studies. The TE was first reported simply as a demonstration with two face pictures (Thompson 1980), which did not allow quantification of the TE magnitude. Several TE studies used a rating scale of grotesqueness as a dependent measure (Bartlett and Searcy 1993; Murray et al 2000; Parks et al 1985; Rock 1988; Valentine and Bruce 1985), which is difficult to generalize outside of the face domain. Other studies used a task in which participants judge whether an individual face or a pair of faces is thatcherized or not (Donnelly and Hadwin 2003; Lewis 2001; Rouse et al 2004), which can be impractical because it would require teaching the participants what 'thatcherization' means for each object category. In the present work, we measured the TE in a sequential matching task (figure 2). On each trial, the two stimuli were always from the same object, either both upright or both inverted, with the two stimuli being either identical (both normal or both thatcherized) or different (a normal image versus a thatcherized image). Participants were required to perform a same–different judgment between the two stimuli. In this case, the TE is operationally defined as the extent to which global inversion of the images makes it difficult to detect orientation changes in local parts (upright d' minus inverted d'). This task appears to capture the original TE demonstration well, wherein differences between a normal and thatcherized face are more easily detected when shown upright than inverted. In addition, while thatcherization for faces appears to be obvious regardless of whether one part (usually the eyes) or two parts (both eyes and the mouth) are locally inverted, it is unclear whether the number of manipulated features affects whether a TE can be obtained with objects. Therefore we tested the TE for objects with either one or two parts manipulated (experiment 1). Results of experiment 1 revealed a TE for various categories of non-face objects, confirming that the TE is not unique for faces.

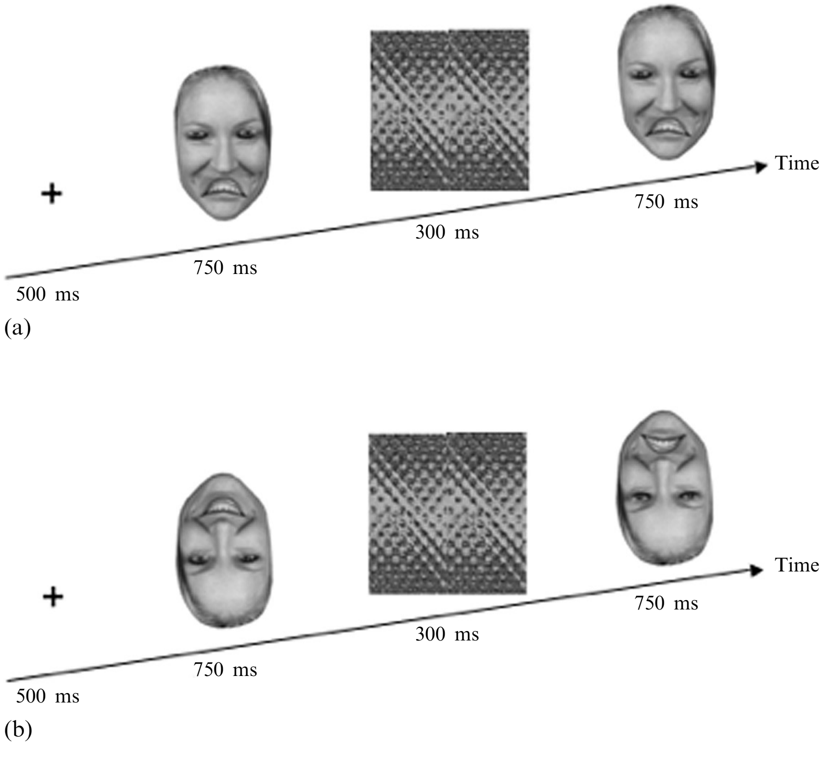

Figure 2. The sequential matching task used in experiment 1. In each trial, participants judged whether two sequentially presented images were the same or different: (a) 'same' trial with human faces with two features locally inverted in the upright condition; (b) 'different' trial in the inverted condition in which the first face is normal and the second face has two features locally inverted.

To compare the TE magnitude across categories is more challenging. Different object categories vary in many dimensions (such as shape, complexity, number of parts and the spatial relationship between the parts) and it is unknown which dimension(s) is(are) relevant to the TE magnitude of a certain object category, or how different factors affect the TE of different object categories differently. For our purposes, however, it is unnecessary to identify all relevant dimensions, because we can measure the overall influence of all relevant dimensions on the TE magnitude with the behavioral performance in the sequential matching task (ie the upright d' and inverted d'). In this study, we carefully matched either the inverted d' (experiment 2) or upright d' (experiment 3) across object categories, and then compared the TE magnitude (upright d' − inverted d') across categories. Given that there is no ceiling or floor effect, a similar performance level (either upright d' or inverted d') suggests that the unknown relevant dimension to the TE provides a similar constraint to the possible TE magnitude for each category, and thus allows us to appropriately compare the TE magnitude across categories. Performance was matched according to the inverted d' (experiment 2) or upright d' (experiment 3), on the basis of different theoretical considerations. Matching performance may be more appropriate for inverted objects, considering that similar processing strategies are available (part-based processing—Farah et al 1998), while matching performance for upright objects may provide better psychometric estimates as the performance was generally better (we thank an anonymous reviewer for pointing this out). Results from both experiments converged to suggest that a TE can be obtained with both faces and non-face objects, with the face TE being the largest (but not exceptionally large) among the tested object categories.

Finally, because we included many object categories, we also explored a few specific hypotheses about factors that may specifically affect faces or words, for which a significant TE is expected on the basis of prior studies and demonstrations (experiment 1). For faces, we tested the importance of grotesqueness as a determining factor. If the difference in grotesqueness between upright and inverted thatcherized faces contributes to the large TE for faces (Murray et al 2000; Parks et al 1985; Rock 1988; Valentine and Bruce 1985), the use of original faces that are already grimacing should reduce the TE, because both normal and thatcherized faces would now contain grotesque features. With regards to words, we explored how the TE is influenced by lexicality and word frequency. We hypothesized that words should produce a larger TE than non-words, as the global inversion effect may be larger for more familiar letter sequences. Similarly, with the conjecture that familiarity may be important to the TE, we expected a larger TE for high-frequency words than for low-frequency words.

2 Experiment 1[†]

2.1 *Methods*

2.1.1 *Participants.* Twenty-one undergraduate students from Vanderbilt University (fifteen females, six males) were recruited for course credit or cash payment. All reported normal or corrected-to-normal vision and gave informed consent according to the guidelines of the institutional review board of Vanderbilt University.

2.1.2 *Stimuli and design.* The experiment was conducted on Power Macs G4 with Matlab (Natick, MA) with the Psychophysics Toolbox extension (Brainard 1997; Pelli 1997). Stimuli consisted of ten objects in each of twelve categories, including faces (adult faces, baby faces, grimacing faces, animal faces), objects (buildings, cars), scenes (close-up scenes, scenes), and letter strings with frequency and lexical status manipulated (high-frequency words, low-frequency words, high-frequency non-words, low-frequency non-words) (figure 1). Grimacing faces were adult faces with grimacing expressions. Eight-letter words were selected from the MRC psycholinguistic database (Coltheart 1981), in which high-frequency (HF) words had a frequency between 204 and 392, and low-frequency (LF) words had a frequency of 1. Non-words were generated from each of these chosen words by switching the positions of internal adjacent letters (2nd and 3rd, 4th and 5th, etc) with the first and last letters kept in the

---

[†] Note: Experiment 1 is from the honors thesis of Elyssa Twedt.

original positions. This was done to keep the global envelope of the words and non-words similar. Other stimuli were collected from various public-access Internet sites or image banks. Letter strings were shown in black on white using lowercase letters in the Arial font and shown at about 2.1 deg × 5.24 deg of visual angle, while other stimuli were shown at about 5.24 deg × 5.24 deg of visual angle in color. Thatcherization was performed by locally inverting features with Adobe Photoshop 7.0 with 180° rotation of local parts of each object. The transformation was local (eg the left eye remained on the left) without disruption of the global shape of each object. In letter strings, the letters transformed were randomly selected after excluding the first and the last letters and letters that have top – down symmetry (eg u versus n, p versus d). Objects had either one or two local parts rotated. Within each category, the same parts were rotated across all object exemplars, although the objects within the scenes were more variable in identity and location, and windows on houses were more variable in their location (table 1). Inverted images (either normal or thatcherized) were created by 180° rotation of the whole images.

Table 1. The summary of the locally inverted parts for various object categories in the three experiments. In experiment 3, 3 local parts were manipulated for faces and animal faces and 2 local parts for the rest of the objects (indicated with the asterisk).

Category	Experiment		
	1 (2 parts)	2 (3 parts)	3 (2 – 3 parts*)
human faces/ animal faces	both eyes; mouth	both eyes; nose; mouth	both eyes; nose; mouth
buildings/houses	two windows or doors	three windows or doors	two windows or doors
CUscenes/scenes	two objects within the scenes	three objects within the scenes	two objects within the scenes
cars	headrest; lights & license plate	headrest; both lights; license plate	both lights; license plate
words	two letters (not 1st or last)	–	–
bikes	–	paddle and two wheels	–

A sequential matching paradigm was used. In each trial, a pair of images based on the same original object was presented, with both images upright or both inverted (figure 2). In each trial, a fixation cross was presented for 500 ms at the center of the screen, followed by the first image for 750 ms, a mask for 300 ms, the second image for 750 ms, and a blank screen for 2250 ms (figure 2). Participants were asked to make a same – different judgment by key press (1 for same trials, and 2 for different trials) on the number pad of the keyboard, and told to respond within 3 s after the onset of the second image. Both accuracy and speed were emphasized and a tone was given as feedback for incorrect or no response trials. There were a total of 1200 trials, with 100 trials for each category. The order of trials for all conditions was randomized. The whole experiment lasted 75 min. Sensitivity (d') and response time (RT) were measured for each condition. $\Delta d'$ (upright d' minus inverted d' for each category) served as our measure of the TE.

Three factors were manipulated within subjects: orientation (upright/inverted), number of thatcherized parts (1/2), and category. Half of the trials were ‘same’ (both normal, both with one part thatcherized, or both with two parts thatcherized). The other half of the trials were ‘different’ (a normal image paired with a thatcherized image, with either one or two parts thatcherized), with the order of the two images counterbalanced. As ‘same’ trials with two normal images could not be categorized into the condition of one or two thatcherized parts, the performance for the 'same' trials for each level of transformation was averaged across 'same' trials with both normal images and images in that level of transformation.

2.2 *Results and discussion*

For each category, the TE was quantified by $\Delta d'$, that was subjected to a 12×2 within-subjects ANOVA (category × number of thatcherized parts) (consideration of RT showed no speed–accuracy trade-off). A significant effect of category was obtained ($F_{11,220} = 19.0$, $p \leqslant 0.0001$) as well as an interaction between category and number of thatcherized parts ($F_{11,220} = 2.67$, $p = 0.0031$) (figure 3). More importantly, planned one-sample t-tests ($\text{df} = 20$) were performed to see whether $\Delta d'$ was reliably different from zero, ie whether a significant TE could be obtained, for each category at each number of thatcherized parts (Bonferroni corrected for 12 contrasts, for 1 part thatcherized and 2 parts thatcherized separately). Results revealed a significant TE for all categories of faces and letter strings, either with one or with two parts inverted. In addition, when two parts were inverted, buildings also led to a significant TE. The two categories of scenes did not produce a significant TE at either number of thatcherized parts. These results confirm that the TE is not exclusive to faces, since it can be obtained at least with letter strings and buildings (Boutsen et al 2006; Parks 1983; Rock 1988; Rouse et al 2004).

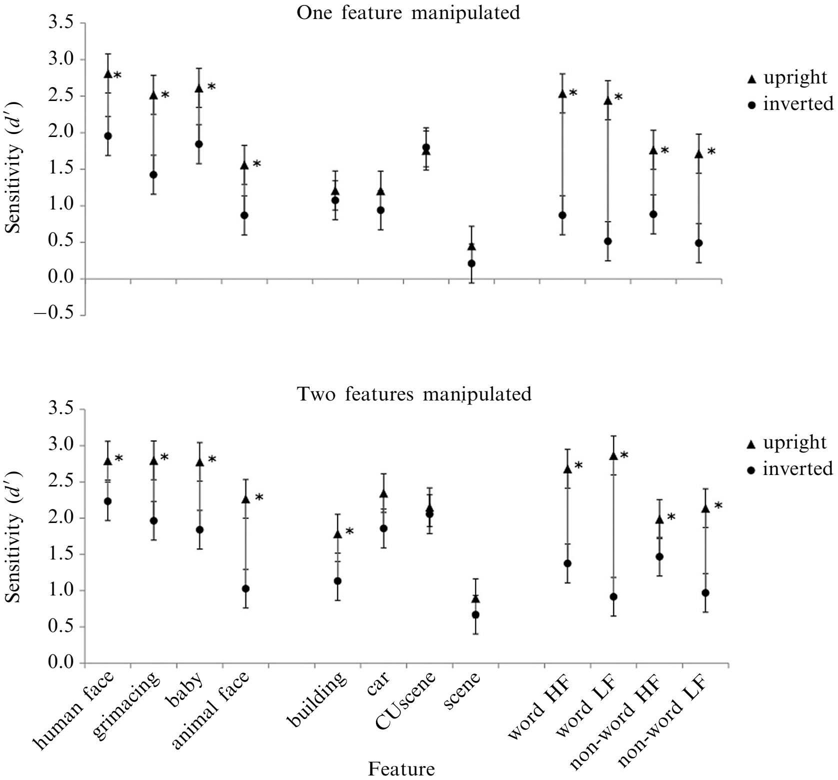

Figure 3. Sensitivity for upright and inverted conditions with the twelve categories of objects in experiment 1 when (a) one or (b) two features were manipulated. The TE was found with all categories of faces and letter strings, and also with buildings when two features were manipulated. Asterisks indicate significant $\Delta d'$ (upright d' – inverted d'), $p < 0.05$. Error bars show the 95% confidence intervals of the 12×2 (category × level of transformation) interaction effects.

2.2.1 *The magnitude of the TE for faces versus non-face objects.* How does the magnitude of the TE for faces compare with that obtained for non-face objects? It is apparent from figure 3 that the TE magnitude varies greatly across object categories. However, across faces and non-face objects, the upright and inverted d' ranged from 0.67 to 2.80 and from 0.44 to 2.09, respectively (figure 3). The large range of upright or inverted d' suggests that detecting local orientation changes is much more difficult for some categories than others. This potentially provided unequal constraints for the TE magnitude across categories (eg a category with an upright $d' = 3$ may have a TE ranging from 0 to 3, while a category with an upright $d' = 0.5$ will have constrained a TE to be equal or smaller than 0.5). Therefore different TE magnitudes can be hard to interpret when overall performance is very different. In experiments 2 and 3, performance on inverted or upright objects was matched, respectively, for a more appropriate comparison of the TE magnitude across face and non-face categories.

2.2.2 *The effect of grotesqueness for faces.* A separate 4 × 2 ANOVA (category × number of thatcherized parts) was conducted on $\Delta d'$ for the four face categories. The interaction between category and number of thatcherized parts was significant ($F_{3,60} = 4.96$, $p = 0.004$). Scheffé tests indicated that, only for animal faces, inverting two parts rather than one produced a larger TE ($p = 0.032$). Considering that these categories had both similar upright d' (Scheffé tests, all ps > 0.18) and processing strategies, our results suggest that the magnitude of TEs for all face categories was similar (except animal faces with two part changes).

These results do not support the idea that a difference in grotesqueness between upright and inverted TE faces contributes to the face TE, as grimacing faces, for which the diagnosticity of a grotesque expression for matching judgments should be at least reduced, did not lead to a smaller TE than that for normal faces.

2.2.3 *The effect of lexicality and frequency for strings.* A 2 × 2 × 2 ANOVA (lexicality × frequency × number of thatcherized parts) was conducted on $\Delta d'$. A main effect of lexicality was obtained ($F_{1,20} = 36.8$, $p \leqslant 0.0001$), with a larger TE for words than non-words. A main effect of frequency was significant ($F_{1,20} = 19.6$, $p = 0.0003$), with a larger TE for LF than HF words. There was also a significant effect of number of thatcherized parts ($F_{1,20} = 5.25$, $p = 0.033$) in which inverting one part produced a larger TE than inverting two parts. No interaction reached significance (the largest $F_{1,20} = 2.99$ and the associated $p = 0.10$).

These main effects suggest that the magnitude of the TE can be influenced by factors specific to the characteristics of a certain object category (letter strings in this case). Detection of locally inverted letters was enhanced by lexicality when the words were presented upright (figure 3). It may be related to the word superiority effect (Reicher 1969), in which letters are identified faster in the context of a word than a non-word. Indeed, some inverted words may resemble non-words on brief presentation (in particular ~27% of the letters in our stimuli become other valid letters once inverted; eg b becomes q), which might explain why lexicality did not facilitate performance when words were inverted.

The finding that HF words had a smaller TE than LF words was inconsistent with our prediction based on word familiarity. The effect of word frequency did not interact with the lexicality of the strings, suggesting that this effect may be related to sub-lexical information that was preserved even after our scrambling manipulation for constructing non-words (only internal letters were scrambled to preserve the global envelope of the string). The fact that non-words were presented randomized with the original words could also have led to some learning effects whereby some non-words may have been encoded as a transformed version of a recently presented word. We did not set out to explore the role of reading-specific mechanisms in the TE, but this could be done with a more systematic and parametric manipulation of the frequency and semantic information of the strings.

Manipulation of the number of thatcherized parts produced contrasting effects for letter strings compared to other objects. For letter strings, locally inverting more parts reduced the TE, while locally inverting two parts produced a larger TE than inverting one part for buildings and animal faces (see above). Local inversion of parts may have affected the external contour of letter strings, which does not apply to other objects. For letter strings, the orientation of each letter combines to construct the global orientation of the word. Therefore more locally inverted letters may directly weaken cues for the global orientation of the word (especially given that almost a third of the letters were valid in both orientations), resulting in a smaller performance difference between upright and inverted words, ie a smaller TE. In contrast, the external contour of common objects or faces, which remains unaffected by local part inversion, provides a more stable frame of reference. Future work could test this account by including additional cues for global orientation, for instance underlining the strings or using a capitalized word.

In sum, our results demonstrate that the magnitude of the TE can be influenced by factors specific to a certain object category, such as lexicality and word frequency for letter strings, factors that are independent from an object's similarity to faces (Parks 1983; Rock 1988).

3 Experiment 2

In experiment 1, we demonstrated significant TEs across a wide range of stimuli. However, it was difficult to interpret the TE magnitudes because the performance was highly variable across faces and non-face objects. The goal of experiment 2 was to match performance across categories based on inverted performance, ie under conditions where theory suggests all categories are processed with the same part-based mechanisms (Farah et al 1998; Leder and Carbon 2006).

In experiment 1, the performance for inverted faces was better than that for inverted non-face stimuli. Therefore several changes were introduced in an attempt to make the inverted performance more comparable across categories. First, all local inversions for faces used a vertical flip instead of a 180° rotation, such that subtle changes in gaze directions could no longer provide any local cues. Second, only one part was changed for each object, and the transformed part was selected randomly from the three parts for each category (table 1), so that the difficulty in detecting the transformed part was more comparable across categories. Finally, stimulus complexity across categories was made more comparable by reducing the number of details and parts for the most complex non-face objects (thereby also reducing the number of parts that may compete for attention). The selection of various categories in this experiment was based on our ability to implement these modifications (for instance, some objects are difficult to simplify and others have too few parts), and letter strings were not included as we focused on the TE for stimuli with an unambiguous global orientation.

3.1 *Method*

3.1.1 *Participants.* Participants were twenty-seven undergraduates from Vanderbilt University (twelve males and fifteen females) recruited for course credit or cash payment. All reported normal or corrected-to-normal vision and gave informed consent according to the guidelines of the institutional review board of Vanderbilt University.

3.1.2 *Stimuli and design.* Six object categories were tested: adult faces, animal faces (eg of cats, lions, tigers, chimpanzees, etc), houses, cars, bikes, and close-up scenes. Twenty-four images were used for each category, collected from various public Internet or image-bank sources. Each object was shown at about 5.24 deg × 5.24 deg of visual angle in color.

To achieve more comparable performance across categories, the thatcherized images were created differently from experiment 1 in the following ways. First, thatcherization was done with Adobe Photoshop 7.0 by vertically flipping a local part of each object (instead of plane rotation) to avoid eye gaze changes when manipulating eyes in faces, which might have given additional hints for detecting local changes for faces in experiment 1. Second, only three possible parts could be locally inverted (eg eyes, nose, and mouth for faces; three windows or doors for houses; headlights, license plate, and headrest for cars, etc—table 1) and objects were simplified by eliminating other parts or extra details with Photoshop. Third, only one part was inverted for each image and this was made unpredictable by randomizing the trial order. Finally, to reduce the variability in close-up scenes, images in this category were created by taking pictures of oriented objects (such as a cup, a plate, a water bottle, a stapler, etc) in a group of three on the same background, with the number of total appearances, positions, and combinations of the objects counterbalanced.

Two factors were manipulated within subjects: orientation (upright/inverted), and category. The same sequential matching paradigm and procedure as in experiment 1 was used. Participants were required to press ‘Z’ for same trials and ‘M’ for different trials. For each thatcherized image, the three possible parts of each object were manipulated for a roughly equal number of times in each condition. There was a total of 576 trials, with 96 trials for each category.

3.2 *Results and discussion*

Six participants were excluded from analyses as their average d' across categories was lower than 0.9. First, we evaluated whether the inverted d' was similar across categories. As shown in figure 4, the values of the inverted d' across categories were much more comparable (range: 0.63 – 1.05, excluding scenes) than in experiment 1 (range: 0.21 – 2.23). One-way ANOVA revealed a main effect of category on inverted d' ($F_{5,100} = 11.668$, $p \leqslant 0.0001$). Scheffé tests revealed that the inverted d' for scenes was significantly higher than for all other categories (all ps < 0.04), while those for faces, animal faces, bikes, cars, and houses were not statistically different (all ps > 0.15). Thus, we restricted our comparison of the magnitude of the TE to these five categories (excluding scenes).

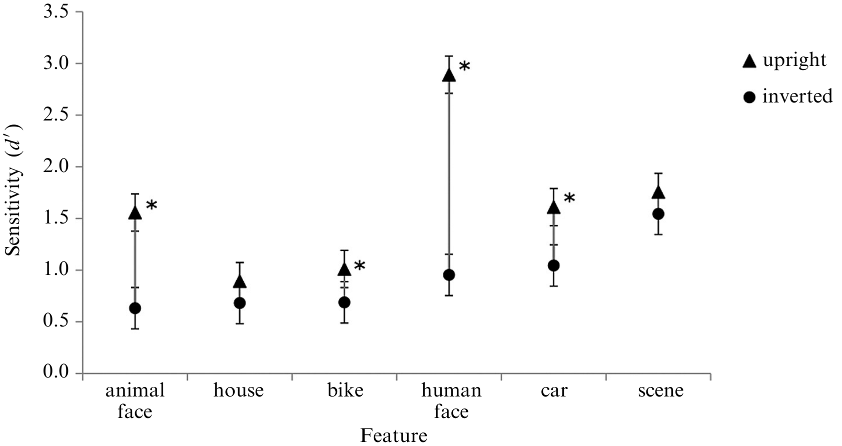

Figure 4. Sensitivity (d') for upright and inverted objects in experiment 2. The TE was found with animal faces, bikes, human faces, and cars. Asterisks indicate a significant difference between upright and inverted d' ($p < 0.05$). Error bars show the 95% confidence intervals of the one-way ANOVA for category on either upright objects or inverted objects only.

To test the significance of TEs, planned one-sample *t*-tests (df = 20) were performed on $\Delta d'$ for the five object categories to see if the TEs were reliably different from zero. We also computed Cohen's *d* for each significant TE effect: four categories yielded significant TEs: faces ($t = 15.2$, $p < 0.05$; $d = 3.11$), cars ($t = 3.12$, $p < 0.05$; $d = 0.76$), animal faces ($t = 5.96$, $p < 0.05$; $d = 1.30$), and bikes ($t = 2.21$, $p < 0.05$; $d = 0.45$). The magnitude of TEs was different across the five categories, supported by a significant one-way ANOVA on $\Delta d'$ for category ($F_{5,100} = 24.1$, $p < 0.0001$). Scheffé tests revealed that the TE for faces was significantly larger than for any other object categories (all *p*s < 0.001). TE for animal faces was also larger than for houses ($p = 0.022$), while those for houses, bikes, and cars were similar (all *p*s > 0.6).

The absence of a TE for houses is consistent with experiment 1 in that no TE was found when buildings had only one part changed. As in experiment 1, scenes did not produce a TE, even though there is clearly no ceiling or floor effect in this condition. In other words, participants could find local changes in scenes, but their performance was not affected by inversion.

In sum, a significant TE was found for faces, cars, animal faces, and bikes. Among the five object categories with matched inverted performance, the magnitude of the TE varied. Faces produced the largest TE, while animal faces produced a TE that was larger than that for houses. Bikes and cars also produced a small but significant TE.

4 Experiment 3

The goal of this experiment was to match performance across categories for upright objects. This was somewhat more difficult: as can be appreciated by inspecting figure 3, faces and objects are generally more similar in inverted than in upright performance. We tried to increase the difficulty of detecting local changes in faces by using neutral faces to reduce the salient change produced by the inversion of smiles. To improve upright performance for non-face objects, we manipulated only two local parts for non-face object categories while keeping three local parts for faces and animal faces (see section 4.1 for details).

4.1 *Method*

4.1.1 *Participants.* Participants were twenty-six undergraduates from Vanderbilt University (nine males and seventeen females) recruited for course credits. None of them participated in either of the previous experiments. All reported normal or corrected-to-normal vision and gave informed consent according to the guidelines of the institutional review board of Vanderbilt University.

4.1.2 *Stimuli and design.* Five stimulus categories were tested: adult faces, animal faces, houses, cars, and scenes. Twenty-four images were used for each category, mostly edited from images used in experiment 2, and the remaining ones were collected from various public Internet or image-bank sources. Each object was shown at about 5.24 deg × 5.24 deg of visual angle in color. The thatcherized images were created differently from experiment 2 in the following ways: to lower upright performance of faces, we replaced laughing or smiling faces with neutral faces so that the thatcherization of the mouths became less obvious. Also, any specularity in the eyes was removed in Photoshop 7.0, such that it could no longer serve as a possible cue for the local orientation of the eyes. To improve upright performance for other object categories, the following measures were used. First, animal faces were further simplified by applying an identical circular outline to all animal faces, so that the different shapes of the external contour and different sizes of the ears would no longer attract attention away from the local face parts. Second, only one of two possible parts could be locally inverted for all non-face objects, while either the eyes, the mouth, or the nose was inverted for faces. For cars, the headrest area was blackened so that it was identical for all images.

For houses, the extra windows or doors were removed. For scenes, only two objects were inserted in an identical background in each image with Adobe Photoshop 7.0 (table 1).

Two factors were manipulated within subjects: orientation (upright/inverted), and category. The same sequential matching paradigm and procedure as in experiment 2 were used.

4.2 *Results and discussion*

First, we evaluated whether the upright d' across categories was similar. The upright d' for faces, scenes, and cars (2.52, 2.48, and 1.94, respectively) was higher than that for houses and animal faces (1.14 and 1.31, respectively). To further attempt to match upright performance across categories, an item analysis was performed. First, the average accuracy for each stimulus in each object category was calculated. Second, the average accuracy for individual stimuli within each category was ranked. The range of upright d' was narrowed down by removing the eight faces that produced the highest average upright performance as well as the eight stimuli that produced the worst average upright performance for the other four categories. Trials for the remaining sixteen stimuli from each category were included in the following analysis. In addition, all data for three participants with the best upright d' for scenes and for three participants with the lowest upright d' for animal faces were removed. Note that this item analysis was based on upright performances, and did not constrain the inverted performances and hence did not affect the magnitudes of the TE across categories (give that there were no ceiling or floor effects). Since we removed the stimuli or participants that contributed to the extreme values, this measure worked against the possibility of obtaining significant results as the d' across categories would become more similar than before.

As shown in figure 5, the upright d' values across categories were more comparable (range: 1.23 – 2.51) than in experiment 1 (range: 0.67 – 2.80) or experiment 2 (range: 0.89 – 2.89). Although there was a difference between the upright d' across categories ($F_{4,76} = 15.9$, $p \leqslant 0.0001$), Scheffé tests revealed that the values of the upright d' were similar between houses and animal faces, between animal faces, cars, and faces, and between cars, faces, and scenes (all ps > 0.09). The upright d' for houses was significantly lower than for cars, faces, and scenes (all ps < 0.001), and the d' for animal faces was lower than for scenes ($p = 0.002$). This allows us to compare the magnitude of the TE among object categories with comparable upright d'.

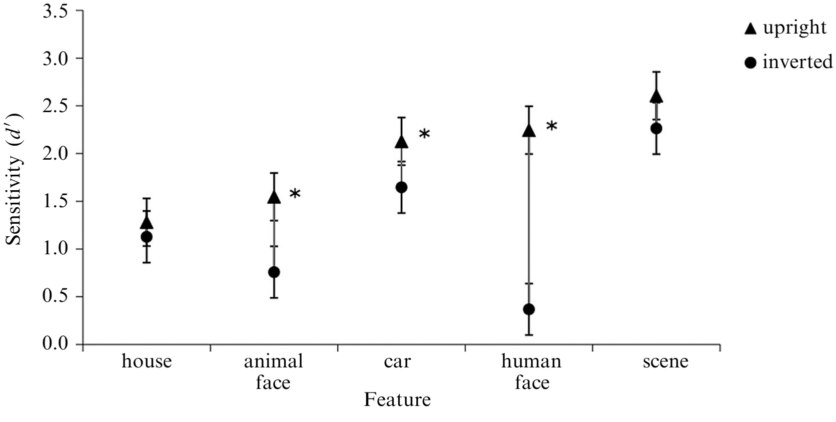

Figure 5. Sensitivity (d') for upright and inverted objects in experiment 3. The TE was found with animal faces, cars, and human faces. Asterisks indicate a significant difference between upright and inverted d' ($p < 0.05$). Error bars show the 95% confidence intervals of the one-way ANOVA for category on either upright objects or inverted objects only.

Planned one-sample t-tests on $\Delta d'$ (df $= 19$) indicated that a significant TE (reliably different from zero) was found for three object categories: faces ($t = 10.9$, $p < 0.05$; Cohen's $d = 2.38$), animal faces ($t = 6.40$, $p < 0.05$; $d = 1.28$), and cars ($t = 3.02$, $p < 0.05$; $d = 0.67$). A one-way ANOVA on $\Delta d'$ for category revealed that the magnitude of TEs differed across categories ($F_{4,76} = 17.7$, $p \leqslant 0.0001$). Scheffé tests revealed that the TE was significantly larger for human faces than for animal faces, cars, and scenes (all ps < 0.015), and larger for animal faces than for houses ($p = 0.009$), while the TE was similar for animal faces and cars ($p > 0.6$). Therefore, with similar performance for upright judgment, faces still yielded the largest TE compared with other objects.

As in experiment 2, a significant TE was not found for houses or scenes when only one part was thatcherized. Note that we successfully equated the upright performance between faces and scenes (Scheffé tests, $p = 0.7$), but a significant TE was only found for faces. The absence of TE for scenes across experiments does not appear to be due to baseline differences.

In summary, the TE was significant for faces, animal faces, and cars. Among the object categories with comparable upright performances, faces led to largest TE, and the TE for animal faces was larger than houses. Therefore, independently of whether performance was equated for inverted (experiment 2) or upright (experiment 3) judgments, we found a larger TE for faces than for non-face objects.

4.2.1 *Is the TE specific for faces?* As demonstrated across our three experiments, the TE can be obtained with both faces and at least some non-face stimuli (eg letter strings, buildings, cars, and bikes). While the TE is not unique to faces, does the face TE stand as an outlier when the effect is measured across a wide range of object categories?

To address this question, we computed the average TE magnitudes within each of seven object categories in our three experiments (faces, animal faces, letter strings, cars, scenes, houses, and bikes). Numerically, the magnitude of the face TE is within the range of TEs obtained with the tested object categories, and is on average smaller than the TE for letter strings (table 2). The effect size of the TE (Cohen's d) resulted in the same pattern (table 2). In addition, we examined the relationship between the averaged TE and the averaged upright or inverted performance within categories. Results revealed that the averaged TE was predicted by the averaged upright d' but not by the averaged inverted d' (upright: $r = 0.804$, $p_5 = 0.029$; inverted: $r = -0.02$, ns; figure 6). In both cases, removing the human and animal face conditions produced similar trends (upright: $r = 0.795$, $p_3 = 0.11$; inverted: $r = -0.27$, ns; figure 6). These results suggest that the magnitude of the TE can be predicted by the ability of observers to detect local changes in upright images, and importantly, the same correlation holds even within non-face object categories.

Table 2. The averaged TE magnitude, upright d', inverted d', and Cohen's d for the tested object categories across three experiments. It shows that the face TE is well within the range of obtained TE among other objects.

Category	TE magnitude	Upright d'	Inverted d'	Cohen's d
scene	0.158	1.59	1.43	0.21
house	0.27	1.28	1.01	0.37
bike	0.323	1.01	0.688	0.45
car	0.48	1.84	1.36	0.58
animal face	0.956	1.78	0.827	1.20
human face	1.1	2.68	1.58	1.53
letter string	1.33	2.27	0.936	1.75

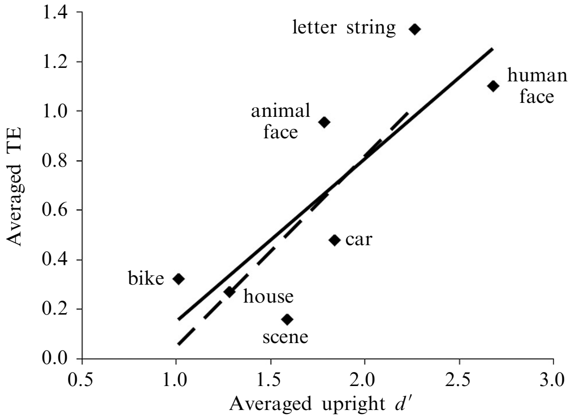

Figure 6. The correlation between the averaged TE and the averaged upright d' across the categories tested in our three experiments. The solid line shows the correlation when all object categories were included. The dashed line refers to the correlation within non-face categories (with human faces and animal faces excluded).

In sum, while faces produce a larger TE than most other object categories, the face TE was not exceptionally large compared to other objects and can be predicted in a similar manner as for non-face objects. While this analysis could be limited by possible ceiling or floor effects that may have restricted the magnitude of the TE estimates (experiment 1), we did not observe evidence supporting the face specificity of the TE.

5 General discussion

We investigated the specificity of the TE for faces first by asking whether the phenomenon is unique to faces, and second by comparing the magnitude of the TE across faces and non-face objects. Using a sequential matching paradigm, we demonstrated that TEs can be obtained for a range of non-face objects including cars, buildings, letter strings, and bikes (experiments 1 – 3). To compare the magnitude of the TE between faces and non-face objects, we matched categories according to matching performance for inverted (experiment 2) or upright (experiment 3) judgments. The magnitude of the TE varied among object categories in both experiments. Across experiments, the magnitude of the TE for faces are larger than for other objects (except LF words, experiment 1), the TE for animal faces was larger than for houses, and no TE was found for scenes.

Note that it is not our goal to make specific claims about the order of the TE magnitudes across categories, nor to uncover specific dimensions modulating the TE magnitudes within or across categories. Our results suggest that the TE magnitude may be dependent on many factors, including stimulus complexity, which parts are transformed, the extent to which observers can predict which parts are transformed (manipulated differently in experiments 1 – 3), etc. For instance, a TE is obtained for buildings and not for cars in experiment 1, whereas a TE is obtained for cars and not for houses in experiment 3. The TE magnitude can also be modulated by factors specific to a certain object category (eg lexicality for letter strings). Nonetheless, one interesting null result in this study is the lack of TE for scenes (the only category that never produces a reliable TE in our three experiments), regardless of whether local change detection was the most difficult (experiment 1), in the middle range of performance (experiment 2), or the easiest across categories (experiment 3). A change-detection study (Varakin and Levin 2008) with scenes using a similar paradigm showed that jumbling the scene interfered with the detection of a change (which could be the deletion or addition of a feature, or an inversion of a feature similar to our study), while inversion of the scene did not (see also Kelley et al 2003; Shore and Klein 2000). These results suggest that there may be a qualitative difference in the spatial relations between parts of an object and those between objects within a scene, such that detecting local inversion of parts is affected by the global orientation of objects but not by the orientation of scenes.

5.1 *The TE is larger for, but not specific to, faces*

Our results consistently revealed a larger TE for faces than for most other object categories even with performance matched, nonetheless without supporting the face specificity of the TE. First, the TE can be found with a range of non-face objects. Second, the magnitude of the face TE falls within the range of TEs from all the object categories we sampled in our experiments (table 2). In other words, the face TE does not appear to be an outlier in the distribution of all possible TEs. Third, the magnitude of the TE can be predicted by the ability to detect local part inversion in upright images regardless of whether the object is a face or not. The evidence suggests that the difference between the TEs for faces and non-face objects may be quantitative, not qualitative, in nature, with a gradient of TE magnitude across object categories.

Indeed, this argument is similar to that for the inversion effect itself. An inversion effect, ie better recognition performance for upright than inverted objects, can be found for a range of face and non-face objects (Palmer et al 1981; Tarr and Pinker 1989; Yin 1969), and the magnitude of the inversion effect varies as a matter of degree across categories (eg Yin 1969). Some authors argue that the larger inversion effects for faces than for non-face objects support face-specific processing mechanisms (eg Robbins and McKone 2007). But others propose that, since the effect is common across object categories and varies in magnitude between faces and non-face objects just like that between non-face object categories, it is not face-specific (eg Ashworth III et al 2007).

It is important to acknowledge the lack of consensus over the interpretation of effects that are not unique to faces but are generally larger for faces. Apart from the TE, two effects that were still argued as being unique to faces (the alignment effect using composite images, and the sensitivity to combinations of spatial frequency and orientation information) were recently obtained with non-face objects (Williams et al, 2009; Wong et al 2009). To account for these effects (either from a face-specific or object-general view), one must do a better job not only in explaining why the effects are larger with faces than with objects, but also why they are found in objects in the first place and can vary significantly among non-face categories (which cannot be explained by mechanisms that are unique to faces, for instance Biederman and Kalocsai 1997). To achieve this, it is important to compare faces to more than one non-face domain, since comparing faces with only one other category may lead to biased conclusions (Gauthier and Nelson 2001), and will not help to account for the variability obtained among non-face categories. We hope that other authors will be encouraged to contrast faces to several other categories in future studies such that a meta-analytical procedure similar to this can provide an even more powerful context to interpret this phenomenon.

5.2 *Factors affecting the TE magnitude*

According to the present results, many factors appear to influence the magnitude of the TE. Some factors may be domain-general. For example, a canonical orientation may be essential (but perhaps not sufficient) for objects to produce a TE such that inversion impairs processing. In addition, the ability to detect local inversion of parts is likely dependent on the extent that configural information across object parts is encoded for a category. Face processing is thought to be particularly sensitive to configural changes (Tanaka and Sengco 1997; Young et al 1987), but part configuration is also important in object perception. For instance, an advantage for the detection of changes in part configuration over other types of changes (such as shape changes) has been reported in a change-detection task with novel objects (Favelle et al 2006). Furthermore, lighting cues (shape-from-shading), which affect both face and object perception, also modulate the magnitude of the TE (Talati et al 2010).

Perhaps most relevant to the large TE for faces, experience with the objects is likely another factor influencing the TE magnitude (Carbon et al 2007), as both sensitivity to orientation and reliance on configural information can be enhanced with experience. For example, experts at individuating objects (eg car experts) demonstrate a larger inversion effect than novices (eg Curby and Gauthier 2007). Also, orientation-specific sensitivity to configural changes can increase with expertise with an object category (Busey and Vanderkolk 2005; Gauthier and Tarr 1997, 2002; Gauthier et al 1998, 2002; Hayward et al 2008; Wong et al 2009). Although we did not measure or manipulate perceptual expertise directly here, some of the largest TEs were found with faces and letters, the two domains for which our undergraduate participants presumably had the most expertise.

Other factors may be more domain specific, which is obviously the case for letter strings, with lexicality and word frequency. For faces, the expressive content carried by a grotesque thatcherized image, something that has no obvious equivalence in objects, is likewise a face-specific factor by definition. Such factors cannot be the basis of comparison across categories. Grotesqueness occupies a central place in discussion and designs concerning the TE (Murray et al 2000; Parks et al 1985; Rock 1988; Talati et al 2010; Valentine and Bruce 1985) but its importance is generally assumed, not tested. However, we found that the magnitude of the face TE was similar even in conditions where the difference in grotesqueness between upright and inverted thatcherized faces was largely reduced. In other words, although grotesque facial expression is sometimes considered the essence of the face TE, we at least did not observe evidence supporting this assumption. We conclude that the contribution of this factor to the face TE remains to be demonstrated.

In conclusion, the TE is not unique to faces or to domains of expertise such as faces and words. Despite the influence of several factors on the TE, some of which may only apply to a restricted number of domains, our results are consistent with the idea that the magnitude of the TE for any category generally depends on our configural skills with the canonical orientation of objects from this category. This hypothesis needs to be tested more directly in studies that experimentally manipulate expertise with non-face objects.

Acknowledgments. This work was funded by NSF grant #SBE-0542013 to the Temporal Dynamics of Learning Center, an NSF Science of Learning Center, and by a grant from James S McDonnell Foundation to the Perceptual Expertise Network. We would like to thank Morgan Webb and Stewart Phillips for their help with creating stimuli and collecting data.

References

Ashworth III A R S, Vuong Q C, Rossion B, Tarr M J, 2007 "Recognizing rotated faces and Greebles: What properties drive the face inversion effect?" *Visual Cognition* **16** 754 – 784

Bartlett J C, Searcy J, 1993 "Inversion and configuration of faces" *Cognitive Psychology* **25** 281 – 316

Bertin E, Bhatt R S, 2004 "The Thatcher illusion and face processing in infancy" *Developmental Science* **7** 431 – 436

Biederman I, Kalocsai P, 1997 "Neurocomputational bases of object and face recognition" *Philosophical Transactions of the Royal Society of London, Section B* **352** 1203 – 1219

Boutsen L, Humphreys G W, 2003 "The effect of inversion on the encoding of normal and 'Thatcherized' faces" *Quarterly Journal of Experimental Psychology A* **56** 955 – 975

Boutsen L, Humphreys G W, Praamstra P, Warbrick T, 2006 "Comparing neural correlates of configural processing in faces and objects: An erp study of the Thatcher illusion" *NeuroImage* **32** 352 – 367

Brainard D H, 1997 "The psychophysics toolbox" *Spatial Vision* **10** 433 – 436

Busey T A, Vanderkolk J R, 2005 "Behavioral and electrophysiological evidence for configural processing in fingerprint experts" *Vision Research* **45** 431 – 448

Carbon C C, Grüter T, Weber J E, Lueschow A, 2007 "Faces as objects of non-expertise: Processing of thatcherised faces in congenital prosopagnosia" *Perception* **36** 1635 – 1645

Coltheart M, 1981 "The MRC psycholinguistic database" *Quarterly Journal of Experimental Psychology A* **33** 497 – 505

Curby K M, Gauthier I, 2007 "A visual short-term memory advantage for faces" *Psychonomic Bulletin & Review* **14** 620 – 628

Donnelly N, Hadwin J A, 2003 "Children's perception of the Thatcher illusion: Evidence for development in configural face processing" *Visual Cognition* **10** 1001 – 1017

Edmonds A J, Lewis M B, 2007 "The effect of rotation on configural encoding in a face-matching task" *Perception* **36** 446 – 460

Farah M J, Wilson K D, Drain M, Tanaka J N, 1998 "What is 'special' about face perception?" *Psychological Review* **105** 482 – 498

Favelle S K, Hayward W G, Bruke D, Palmisano S, 2006 "The configural advantage in object change detection persists across depth rotation" *Perception & Psychophysics* **68** 1254 – 1263

Gauthier I, Nelson C A, 2001 "The development of face expertise" *Current Opinion in Neurobiology* **11** 219 – 224

Gauthier I, Tarr M J, 1997 "Becoming a "Greeble" expert: Exploring mechanisms for face recognition" *Vision Research* **37** 1673 – 1682

Gauthier I, Tarr M J, 2002 "Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior" *Journal of Experimental Psychology: Human Perception and Performance* **28** 431 – 446

Gauthier I, Williams P, Tarr M J, Tanaka J, 1998 "Training 'Greeble' experts: A framework for studying expert object recognition processes" *Vision Research* **38** 2401 – 2428

Hayward W G, Rhodes G, Schwaninger A, 2008 "An own-race advantage for components as well as configurations in face recognition" *Cognition* **106** 1017 – 1027

Kelley T A, Chun M M, Chua K P, 2003 "Effects of scene inversion on change detection of targets matched for visual salience" *Journal of Vision* **3** 1 – 5

Leder H, Carbon C C, 2006 "Face-specific configural processing of relational information" *British Journal of Psychology* **97** 19 – 29

Lewis M B, 2001 "The Lady's not for turning: Rotation of the Thatcher illusion" *Perception* **30** 769 – 774

Lewis M B, Johnston R A, 1997 "The Thatcher illusion as a test of configural disruption" *Perception* **26** 225 – 227

Murray J, Yong E, Rhodes G, 2000 "Revisiting the perception of upside-down faces" *Psychological Science* **11** 492 – 496

Palmer S, Rosch E, Chase P, 1981 "Canonical perspective and the perception of objects", in *Attention and Performance IX* Eds J Long, A Baddeley (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 135 – 151

Parks T E, 1983 "Misprint?" Letter to the editor" *Perception* **12** 88

Parks T E, Coss R G, Coss C S, 1985 "Thatcher and the Cheshire cat: Context and the processing of facial features" *Perception* **14** 747 – 754

Pelli D G, 1997 "The videotoolbox software for visual psychophysics: Transforming numbers into movies" *Spatial Vision* **10** 437 – 442

Reicher G M, 1969 "Perceptual recognition as a function of the meaningfulness of the stimulus material" *Journal of Experimental Psychology* **81** 275 – 280

Robbins R, McKone E, 2007 "No face-like processing for objects-of-expertise in three behavioural tasks" *Cognition* **103** 34 – 79

Rock I, 1988 "On Thompson's inverted-face phenomenon (research note)" *Perception* **17** 815 – 817

Rouse H, Donnelly N, Hadwin J A, Brown T, 2004 "Do children with autism perceive second-order relational features? The case of the Thatcher illusion" *Journal of Child Psychology and Psychiatry* **45** 1246 – 1257

Shore D I, Klein R M, 2000 "The effects of scene inversion on change blindness" *Journal of General Psychology* **127** 27 – 43

Talati Z, Rhodes G, Jeffery L, 2010 "Now you see it, now you don't: Shedding light on the Thatcher illusion" *Psychological Science* **21** 219 – 221

Tanaka J W, Sengco J A, 1997 "Features and their configuration in face recognition" *Memory & Cognition* **25** 583 – 592

Tarr M J, Pinker S, 1989 "Mental rotation and orientation-dependence in shape recognition" *Cognitive Psychology* **21** 233 – 282

Thompson P, 1980 "Margaret Thatcher: A new illusion" *Perception* **9** 483 – 484

Valentine T, Bruce V, 1985 "What's up? The Margaret Thatcher illusion revisited" *Perception* **14** 515 – 516

Varakin D A, Levin D T, 2008 "Scene structure enhances change detection" *Quarterly Journal of Experimental Psychology* **61** 543 – 551

Williams N R, Willenbockel V, Gauthier I, 2009 "Sensitivity to spatial frequency and orientation content is not specific to face perception" *Vision Research* **49** 2353 – 2362

Wong A C-N, Palmeri T, Gauthier I, 2009 "Conditions for face-like expertise with objects: Becoming a Ziggerin expert—but which type?" *Psychological Science* **20** 1108 – 1117

Yin R K, 1969 "Looking at upside-down faces" *Journal of Experimental Psychology* **81** 141 – 145

Young A W, Hellawell D, Hay D, 1987 "Configural information in face perception" *Perception* **16** 747 – 759

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 39 2010

www.perceptionweb.com